mental health digest

...nal Clearinghouse for Mental Health ...rormation, Vol. 4, No. 11, November 1972

mhd

MAYO CLINIC
LIBRARY

NOV 14 1972

ROCHESTER, MINN.

Cross out the word that does not belong with the other three words:

egg rug woman sex

# cover story

**Diagnoses, Classifications, Tests, Scales, and Other Measurements**

**The item on the cover is a kind often used in psychological tests. It also illustrates a perplexing problem in testing, classification, and diagnosis, as shown by three possible answers:**

1. **Cross out "rug." The other three words are related to reproduction.**
2. **Cross out "woman." The other three words are three-letter words.**
3. **Cross out "sex." You can beat an egg, a rug, and a woman.**

The *Mental Health Digest* is a monthly publication designed to reflect the whole spectrum of mental health and to present a broad sampling of scientific subject matter and professional points of view. It contains condensations of material selected from the current mental health literature. The condensations are published with the approval of the authors and publishers for use in the *Mental Health Digest* only. Readers interested in particular digests are urged to consult the journal or other source in which the material first appeared or to ask the authors for reprints.

Material contained in the *Digest* does not necessarily reflect the opinions, official policy, or position of the National Institute of Mental Health.

*Editorial Advisory Board*

Bertram S. Brown, M.D., Chairman

William E. Bunney, M.D.
Harry P. Cain II, Ph.D.
Morris E. Chafetz, M.D.
Robert A. Cohen, M.D., Ph.D.
John C. Eberhart, Ph.D.
James A. Goodman, Ph.D.
James D. Isbister
Thomas F. A. Plaut, Ph.D.
Luther D. Robinson, M.D.
Gian C. Salmoiraghi, M.D., Ph.D.
Claudewell S. Thomas, M.D.
Louis A. Wienckowski, Ph.D.

*Editor:* Lorraine Bouthilet, Ph.D.

# to subscribe

The *Mental Health Digest* is available on subscription from the Superintendent of Documents, Government Printing Office, Washington, D.C. 20402. Subscription price: $3.50 a year; Single copy: 30¢. For foreign mailing, add $1.00. (No additional postage is required for Canada, Mexico, and Central and South American countries except Argentina, Brazil, British Honduras, French Guiana, Guyana, and Surinam.) Please make check or money order payable to the Superintendent of Documents. An annual subscription runs for 12 consecutive issues and begins when orders are processed. However, a subscription may be backdated to include issues already published in the current volume when the issues are available. Issues for back years, when available, are sold at the single copy price. The expiration date is printed on the top line of the address label. Change of address and complaints about nonreceipt of issues should be sent to the Superintendent of Documents, and should include the old address, preferably a mailing label, as well as the new address.

Use of funds for printing this publication approved by the Director of the Bureau of the Budget, May 20, 1968.

# contents

**Mental Health Digest, Vol. 4, No. 11, November 1972**


**1** Psychiatric Distinctions: New and Old Approaches
by Donald Conover
from the *Journal of Health and Social Behavior*

**9** Beyond Diagnosis: An Approach to the Integration of Behavioral Science and Clinical Medicine
by Ian Renwick McWhinney
from the *New England Journal of Medicine*

**13** Validity of an Empirically Derived Phenomenological Typology
by John E. Overall, B. W. Henry, and James R. Markett
from the *Journal of Psychiatric Research*

**21** Syndromes of Psychosis and Nonpsychosis: Factor Analysis of a Systems Analysis
by Raymond D. Martorano and Peter E. Nathan
from the *Journal of Abnormal Psychology*

**24** Studies of Clinical Judgment: Part II. Toward a New Model for Research
by Michael K. Harty
from the *Bulletin of the Menninger Clinic*

**31** Personality and IQ Measures as Predictors of School Achievement
by K. Barton, T. E. Dielman, and R. B. Cattell
from the *Journal of Educational Psychology*

**34** A Nonverbal Semantic Differential
by P. M. Bentler and Allan L. Lavoie
from the *Journal of Verbal Learning and Verbal Behavior*

**contents continued**


37 The Psychodynamics of Examinations
by M. L. Moeller
from the *Zeitschrift fur Psychotherapie und medizinische Psychologie*

43 Transitions in Infant Sensorimotor Development and the Prediction of Childhood IQ
by Robert B. McCall, Pamela Savoy Hogarty, and Nancy Hurlburt
from the *American Psychologist*

50 Comparative Validities of Direct and Indirect Personality Tests
by William A. Scott and Ronald C. Johnson
from the *Journal of Consulting and Clinical Psychology*

57 Getting Information from School and Clinical Psychologists
by Julia R. Vane
from *Professional Psychology*

## mental health and the law

59 Courts, State Hospitals, and the Right to Treatment
by Jonas Robitscher
from the *American Journal of Psychology*

## 65 mental health news digests

# Psychiatric Distinctions: New and Old Approaches

**Donald Conover**

Although it has long been asserted that practical and theoretical problems are associated with making psychiatric distinctions, only recently have these problems received empirical study. One purpose of this paper is to bring together recent studies so as to provide a picture of the strengths and weaknesses of current methods of making psychiatric distinctions.

The paper is organized into three parts. First, the traditional method for making psychiatric distinctions will be discussed. The second part is devoted to curent developments in psychiatric research. The final section deals with projections of curent trends in research.

**Traditional Methods**

In considering psychiatric diagnoses we are concerned with questions of the reliability of distinctions that are made within the mentally ill category. Validity is not dealt with because there are, as yet, no acepted "external independent criteria" against which the validity of psychiatric diagnoses can be measured. The distinction between the mentally ill and the mentally well is not examined because the assumption is made that the individuals being diagnosed are in fact mentally ill.

Criticism of the low reliability of psychiatric diagnosis frequently has been made in the psychiatric literature. The agreement rate between independent psychiatrists on their diagnoses of the same patient has been shown to be about 60 percent in those studies where the procedures most closely approximate the situation of clinical diagnosis. But even in these studies, procedures were used which would probably raise the level of agreement over that in a truly clinical setting. Procedures, such as meetings to discuss diagnostic categories, were used in a study that found an agreement rate of 56 percent for the various subcategories of mental illness.

If these preliminary meetings had not been held, the research would have more closely approximated "field" conditions and, in all likelihood, the rate of agreement would have been lower. When fewer and grosser categories of mental illness are used, however, the agreement rate, as would be expected, increases.

In another study, all the diagnosticians (eight from North Carolina, four from London, and four from Glasgow) were provided with the same information on a number of patients. This procedure eliminated variation between diagnosticians in the first phase of diagnosis, the information-gathering phase. For the evaluatory phase, when it is judged whether or not bits of information about the patient and the patient's behavior are pathogenic or symptomatic, variance between diagnosticians is not reduced below the level that would be expected in clinical practice. Similarly, variance is not reduced in the third phase when the symptoms and clues are organized into diagnostic categories.

Each diagnostician viewed films of 30 U.S. patients being interviewed by U.S. psychiatrists. The diagnosticians also were given the following information on each patient: age, sex, admission status, family history, physical examination and laboratory findings, the Wechsler Adult

Intelligence Scale score, the (uninterpreted) MMPI profile, description of treatment given during the first 2 weeks in hospital, and extracts from notes on the patient's course during this period.

The fact that there were twice as many diagnosticians from North Carolina than from either London or Glasgow would seem to put the North Carolina diagnosticians at a slight disadvantage. The researchers realized this and used a statistic which overcomes the difficulty arising from the unequal number of diagnosticians. Using this equalizing statistic, the following levels of agreement were found: Glasgow 73 percent, London 64 percent, and North Carolina 58 percent.

The evidence thus far shows that while the reliability of psychiatric diagnoses is quite low, it is seen as varying acording to the research procedures and the nationalities of the diagnosticians. We shall now turn to studies that are aimed at seeking reasons for the low reliability of psychiatric diagnosis.

One study used a rather straightforward approach to determine why two psychiatrists, assessing the same patient, arrive at a different diagnosis. Four experienced psychiatrists were used. In the preliminary phase, a number of patients were interviewed separately and in quick succession by two psychiatrists. After the interviews, each lasting about an hour, the psychiatrists compared their diagnoses. When they disagreed, the reasons were determined. A list was then made of these reasons. In the second phase, the list of reasons was referred to in the 40 cases where disagreement occurred, and one reason was selected as primary in each of the cases. Tabulations were then made to determine the frequency of the different reasons.

The findings showed that reasons having to do with "inadequacies of the nosology" accounted for 62.5 percent of the disagreements. Accounting for 32.5 percent of the disagreements was "inconsistencies on the part of the diagnosticians." The most important reason under this second heading was "weighing symptoms differently," which accounted for 17.5 percent of the total number of disagreements. The least important class of reasons was "inconsistencies on the part of the patient," which accounted for 5 percent of the cases.

Conversion reaction and psychophysiological reaction were the two most frequently confused diagnoses. This confusion is due, the researchers feel, to the nosology, which for these diagnoses requires highly inferential distinctions about patients. Another weakness in nosology that makes agreement in diagnosis problematic is unclear criteria.

In 1968, a second edition of the *Diagnostic and Statistical Manual of Mental Disorders* was published by the American Psychiatric Association. Unfortunately, the definitions and criteria are still quite vague and loose. Moreover, an attachment to old terms has hampered some efforts of clarification—e.g., we learn of the apparent paradox that some psychoses are "not psychotic." Regarding psychiatric nomenclature there is, however, one note of cheer: Today, in the United States, there is only one nosology to be confused about, whereas in the past there were several.

In another study, the approach was to isolate and measure the reliability of a single component—i.e., a stereotype—of psychiatric diagnosis. Their "stereotype," when put into our schema, is identified as the third phase of diagnosis wherein symptoms and clues are organized into diagnos-

tic categories. The procedure was to match individual psychiatrists' stereotypes of 20 common psychiatric disorders with "master" stereotypes of these disorders. The master stereotypes were constructed by including in each the four to six most frequently mentioned symptoms or characteristics listed in the individual stereotypes of a particular disorder. They counted as an "agreement" all cases in which the symptoms and characteristics of an individual stereotype occurred more frequently in its nominal "master" counterpart than in any of the 19 other master stereotypes. Using this method, they found agreement rates ranging from 81 to 88 percent.

If we accept 81 to 88 percent agreement as being a "high" rate of agreement (this is questionable since the two items compared—individual and master stereotypes—are related by definition), then the problem of reliability would seem to lie in the other components of diagnosis, that is, in phases one and two. The following research concerns questions of reliability at the earlier phases of diagnosis.

One interesting finding was that British and American psychiatrists, referring to the same case material, differ markedly in the number of times that they regard behavior as symptomatic. If this finding is combined with the finding that the percentage of agreement on diagnoses was considerably higher among the British psychiatrists than among the American psychiatrists—Glasgow 73 percent, London 64 percent, and North Carolina 58 percent—we come to the interesting conclusion that the more symptoms psychiatrists take note of, the harder it is for them to agree on a single diagnosis. The conclusion may be based on rather shaky evidence. Nevertheless, it is a conclusion which, if founded, could have implications regarding the optimal amount of information to be used in making reliable psychiatric diagnoses.

There also were differences between American and British psychiatrists in types of symptoms and in corresponding types of diagnoses. The tendency for British psychiatrists to overdiagnose (relative to American standards) affective disorders and underdiagnose schizophrenia is a finding that is consistent with differences in hospitalization rates in Britain and America for these two diagnoses. The evidence suggests that hospitalization rates for particular diagnostic categories may reflect, in part, the preferences of diagnosticians for certain diagnostic labels. And physicians' preferences may, in turn, reflect national norms as to what is regarded as deviant. British psychiatrists, and perhaps the British in general, may regard fewer acts as symptomatic, and those that they regard as symptomatic are acts that contrast with the normal conservative, reserved mode of British behavior—i.e., elated, excited, and grandiose acts.

One should be cautious in generalizing too quickly. Both studies used films of American patients being interviewed by American psychiatrists. It is possible that the British noted fewer symptoms because they had difficulty understanding the patient's accent. Another possibility is that American pathology, in certain instances, is British normality. All the films were made of Americans whose behavior was regarded by themselves or other Americans as abnormal.

We have talked about symptoms and diagnoses but we have said little about the relation between the two. The closest we have come was in our discussion of stereotypes and in our conjecture regarding the relation be-

tween number of symptoms and reliability of diagnosis. There is a group of studies that does provide some evidence of the existence of a particular qualitative link between symptom and diagnosis. In several studies, it was noted that the symptoms "apathy" and "perceptual distortion" distinguish between diagnoses of psychosis and neurosis. That is, when a comparison is made of the frequency distribution of symptoms, between those psychiatrists who diagnosed a patient as neurotic and those who diagnosed the same patient as psychotic, two results stand out: The psychotic group noted a greater total number of symptoms than the neurotic group and there was a greater frequency of apathy symptoms noted in the psychotic group. This suggests that American clinicians regard a patient as suffering from the more serious class of mental disorders when the patient exhibits many symptoms and when the patient is believed to exhibit a certain kind of symptom—apathy.

Another study differs in that it explores a problem that extends beyond, but includes, the problem of psychiatric diagnosis. The question it seeks to answer is why the proportion of schizophrenic and affective disorders differs in the admission rates of British and American mental hospitals. One possibility is that British and American patients tend to suffer from different forms of psychopathology. Another possibility is that different utilization patterns exist in America and Britain. And a third possibility is that British and American psychiatrists tend to affix different psychiatric labels to patients who are otherwise similar.

The method was to draw two samples of patients, one from a British mental hospital and another from a comparable American mental hospital. Information was collected on both sets of patients using a standardized mental-state schedule and a standard history schedule. The standardized information was then evaluated by a research psychiatrist and a diagnosis was made. These diagnoses were used as controls to compare with the diagnoses made independently by the staffs of the two hospitals.

The findings are quite similar to the ones reported in other studies but the fact that they were arrived at by way of a different research design suggests that the basic conclusion, that British and American psychiatrists each have their own particular "diagnostic bias" for certain patients, seems sound.

### Current Developments

We now direct our attention to recently developed methods for making psychiatric distinctions. The scientists at Biometrics Research have systematized the making of psychiatric distinctions at each major phase of the process. For the information-gathering phase, they have developed interview schedules; for the evaluation phase, rating scales have been produced; and for the organization phase, they have written computer programs. Their work has advanced to the point where it seems almost certain that in the near future a lay interviewer having a moderate amount of training and using an interview schedule, a rating scale, and a computer program will be able to produce a psychiatric diagnosis equal, if not superior, in validity and reliability to a diagnosis made by a psychiatrist using traditional clinical procedures.

A number of instruments for the gathering and evaluation of psychiatric information have been developed. In order to clarify the nature of these instruments, two have been

chosen, the Mental Status Schedule (MSS) and the Psychiatric Evaluation Form R2 (PEF), to be described in greater detail.

*Mental Status Schedule*—The MSS is the first standardized interview designed to assess the major dimensions of mental status in which the content and order of questions are fixed. The MSS measures the severity of psychopathology using either 16 factor-analysis-derived scales, or 23 scales "based on traditional clinical diagnostic considerations." The score for each scale is based on the sum of the number of items in that scale that have been checked "true." As examples, the following are the names of two factor-analysis scales: "Feelings-Concerns" (72 items) and "Retardation-Withdrawal" (15 items); and two traditional clinical scales: "Guilt" (six items) and "Delusional-Grandiosity" (four items). The interview schedule consists of 82 questions, 51 probe questions, and an accompanying inventory of 248 items that the interviewer checks off as true or not depending on the respondent's reply to the questions.

The reliability of the MSS was determined by having two or three psychiatrists present who independently checked items for some 67 interviews with patients in four different psychiatric institutes. The reliability was high, with a mean reliability coefficient of .94 for the 67 patients.

The validity of the MSS was determined by the known-group criterion method and the concurrent method. The mean scores for the following four groups of patients were determined: mental hospital patients, former mental patients, clinic outpatients, and orthopedic hospital patients. As expected, the mental hospital patients had the highest mean score, the orthopedic patients had the lowest mean, and the other two groups were intermediate between the two.

The other validity measure involved comparing the evaluations of severity of illness made by two psychiatrists, one using clinical methods and the other administering the MSS, for 17 formerly hospitalized patients. The patients were independently ranked by the psychiatrists and the correlation was found to be .82.

In sum, the MSS has shown itself to be a reliable, valid, and easy to administer instrument for the assessment of the severity of overall psychopathology. Moreover, it has shown itself to be sensitive to changes in patient psychopathology and to particular areas of behavior. While the MSS does not organize evaluated information into diagnostic categories, it does serve, it seems, the functions of information gathering and evaluation.

*Psychiatric Evaluation Form*—Although the PEF serves the same basic functions as the MSS, it differs in a number of important respects. First, the data used in making the PEF evaluations may be gathered by means of the PEF interview schedule or it may be gathered by other means: unstructured clinical interviews of the patient and/or an informant, case records, or nurses' notes. All data used in MSS evaluations, on the other hand, are gathered directly with the MSS interview. Second, evaluations are made after the data have been gathered, rather than concurrently. Third, the evaluations made in the PEF are not the concrete, easy to make, true-not true decisions regarding a large number of specific items, but are judgments of rather abstract or global items such as "grandiosity," "anxiety," and "social isolations" ranked on six-point scales of severity ranging from none to extreme.

If the interview is used, the proce-

dure in administering the PEF is to ask the patient all the questions, then make ratings on six-point scales for the 19 areas of functioning and the five areas of role impairment. Because the judgments in making PEF evaluations are more complex than those for the MSS, more training is required for the interviewer-rater.

The advantages of the PEF are that it is quicker to administer and it does not require that information be gathered by structured interview—the ratings can be made on the basis of available information. Counterbalancing its advantages is a major disadvantage; it is less reliable and probably less valid than the MSS. The reasons for the lowered reliability and validity stem, it would seem, from the characteristics, listed above, which make the PEF a different instrument from the MSS. That is, error enters from variability in gathering the data, memory bias, and having to make "macro" rather than "micro" judgments. Despite these problems, it is likely that the PEF is more useful or better adapted for certain situations and problems than is the MSS.

*Diagno II*—The most recent work of this group complements their earlier work. They have been trying to simulate clinical judgment, that is, diagnosis making, by the use of computer programs. The advantage of computers is that they do as they are told. Once they have been programed they will repeat precisely the same procedures in carrying out their task, barring mechanical failure. Thus, with computers, the problem of reliability disappears and the researcher can shift his attention to validity—that is, making certain that what the computer does is what he wants done. This group wants the computer to organize a set of evaluated information into the same diagnostic categories that a trained psychiatrist would. Thus far, two programs have been developed for this purpose. We will describe the performance and some of the characteristics of the second and more advanced program, "Diagno II."

The input for Diagno II are 94 scaled (six levels) judgments taken from an instrument called Current and Past Psychopathology Scales (CAPPS), which is a selective combination of the Psychiatric Evaluation Form—Diagnostic Version and the Psychiatric History Schedule. A diagnosis chosen from 44 official APA diagnoses and two unofficial diagnoses—"nonspecific illness with mild symptomatology" and "not ill"—is the output.

The program directs the computer to ask yes-no questions of the input data; for example: "Is disorientation-memory (current state) less than four (on the six-point scale)?" If the answer is yes, then the computer is directed to ask another question, and if the answer is no, the computer asks an alternative question. In this way, the computer is guided along a logical path to a particular diagnosis. The questions and cutoff points used in Diagno II were selected by testing a pool of initial questions against 52 real and 92 hypothetical cases.

The validity of Diagno II was determined by comparing the diagnosis made by Diagno II with diagnoses made by two diagnosticians and a rater for each of 100 real cases and 46 hypothetical cases. The raters filled out a CAPPS for each case, made a diagnosis, then gave their ratings to the computer which made its diagnosis using the Diagno II program. The diagnosticians made their diagnoses from protocols that contained the same information as Diagno II received in the ratings, but in differ-

ent form. The raters had direct contact with the subjects, while the computer and the diagnosticians did not. Also, the raters had a greater amount of information on each subject than did the diagnosticians and the computer.

The findings show that "for real cases, Diagno's diagnoses agree with psychiatrists' diagnoses as well as the psychiatrists agree with each other." While there was higher agreement on the hypothetical cases than on the real cases, the computer diagnoses did not agree with the psychiatrists' on the diagnoses of the hypothetical cases quite as well as the psychiatrists agreed with each other.

In looking at the distribution of the diagnoses for the three types of judges—raters, diagnosticians, and computer—it is interesting to see that the psychiatrists generally, and especially the psychiatrists having the most information about the cases, tend to affix diagnoses suggestive of more severe psychopathology more frequently than does the computer. For example the psychiatrists having the most information about the cases diagnosed 45 percent of the cases as schizophrenic, whereas the computer found 29 percent schizophrenic. At the other end of the continuum, "nonspecific illness" and "not ill," the respective percentages for the two types of judges are 14 and 26. This suggests that while the diagnosis made by the computer often agrees with the one made by the psychiatrists, when disagreements occur, it is more often the psychiatrists who opt for the diagnosis having the less favorable prognosis. The question arises, are the psychiatrists overly pessimistic or is the computer overly optimistic in making diagnoses? The question remains unanswered at this time.

*Stanford University Group*—The last research to be reported here concerns the work of a group at Stanford University. They too are attempting to use a computer to systematize a phase of the diagnostic process. Instead of focusing on the categorization phase, they apply the computer's capabilities to the information-gathering phase.

This group has devised a system whereby video-display computer terminals are set up in an inpatient psychiatric facility. The computer is programed to communicate directly with the patient by means of questions presented on its video display. The questions may be multiple choice or open ended and the choice of question is contingent on the patient's previous responses. For example, if the patient types "yes" in reply to the computer's question, "Have you ever received treatment here before?", the computer asks next, "How long ago?" If the patient's first reply is "no," then the computer asks "Have you received psychiatric treatment elsewhere?" In this manner the patient is asked only logically relevant questions and the computer is able to gather and store a great deal of information on a patient in a short time. Moreover, since the patients are all responding to the same computer program and to questions being asked in the same manner, the information should be highly comparable.

A question regarding this approach to patient interviewing is whether patients want to use it. While the initial results suggest that they do, it remains to be seen whether it will be used voluntarily and consistently by more than a small circle of devotees comparable to the ping-pong or checkers players found on institutional wards. The question seems to come down to whether inmates prefer to communicate with their super-

iors on a personal or impersonal level. It is likely that different patients, or the same patients at different times, prefer different modes of communication, and that the computer terminal will prove helpful in giving patients one more alternative.

From the researcher's point of view, the question of patients' continued preference or acceptance of the computer is less crucial since, in most instances, it would be necessary to gather information on only a few occasions. The main difficulty would be, instead, being certain that the questions are phrased so as to minimize misinterpretation. Although the use of interviewers increases variability, it also serves the function of "catching" misinterpretations—something a computer cannot do.

Despite difficulties and limitations in practice, use of a computer to resolve the information-gathering problem represents a genuine advance in psychiatric research. Instruments and techniques similar to the ones that are being produced will, quite likely, replace the traditional psychiatric diagnosis, at least for researchers, as the preferred means for making psychiatric distinctions. Clinicians will probably follow the lead of researchers and reluctantly put aside many traditional diagnostic practices. As the diagnostic process becomes rationalized, greater empirical understanding will emerge and the diagnostic categories will become more clearcut and meaningful. This will open up greater understanding of etiology and the effects of treatments. The capacity reliably to make psychiatric distinctions is seen as a crucial first step in psychiatric research which is not being taken and which will have far-reaching effects indeed.

---

Condensed with permission from the *Journal of Health and Social Behavior*, 13(2): 167–180, 1972. Copyright 1972 by the American Sociological Association. Reprints of original article from Donald Conover, Ph.D., Programa de Investigaciones en Planeación de Salud, Apartado Aereo 4074, Cali, Colombia

# Beyond Diagnosis: An Approach to the Integration of Behavioral Science and Clinical Medicine

**Ian Renwick McWhinney**

Before we can think in a clear and precise way about any order of phenomena, we have to develop a way of organizing our observations. We do this by classifying the phenomena in ways that enable us to see relations between them and other orders of phenomena. In clinical diagnosis, we categorize diseases and give them names. In the diagnostic method, therefore, medicine has developed a tool by which new knowledge can be applied to the solution of clinical problems.

Our success in applying the biological and physical sciences to medicine is in marked contrast to our failure in applying the behavioral and social sciences. Our problem is that we do not have a ready tool, like the diagnostic method, by which to apply these sciences to the behavior of patients. Lacking any system of classification or taxonomic vocabulary, we have no way of organizing our observations and of communicating our thoughts to others.

The confusion is made worse by our tendency to assume that the diagnostic method, so successful for the utilization of biological knowledge, will be equally successful for the utilization of knowledge from behavioral science. By using the same taxonomic system for clinical and behavioral phenomena, we have made it difficult to separate the two and to study the relations between them. We have no way of classifying behavior that occurs concurrently with an organic illness. Yet every patient exhibits some form of behavior. The fact that this behavior may be considered normal by the physician does not remove the need to describe and classify it.

All that has been said about patient behavior can be said with equal force about the social determinants of this behavior. We do, of course, take social histories from our patients. Too often, however, these are records of isolated facts, with no attempt to relate them to the patient's illness or his behavior.

### A Taxonomy of Patient Behavior

The taxonomy of patient behavior takes the doctor-patient contact as the reference point. For any contact between doctor and patient it should be possible to place the patient's behavior in one of the following categories. The categories are mutually exclusive: A patient should only be placed in one category at any one time.

*Limit of tolerance*—The patient comes because his symptoms are causing pain, discomfort, or disability that have become intolerable. This category covers many straightforward episodes of illness from an attack of influenza to a fractured femur. The capacity of a sympton to cause enough distress for a physician to be consulted depends on many variables, including the patient's social and cultural background, and on his belief that an effective remedy is available. A subcategory comprises patients who come with a problem of living rather than a symptom. Patients come to doctors with unhappiness as well as with illness. Often, unhappiness is expressed as a symptom or symptoms. Sometimes, however, patients present a frank problem that they are no longer able to tolerate. It is important to differentiate problems of liv-

ing from illnesses, since they require different management.

*Limit of anxiety*—The patient comes, not because his symptoms are causing distress, but because of their implications. By definition, anxiety is a component of these episodes. The patient comes because he, or a relative, fears the consequences of his symptoms. Since they depend on a person's knowledge and beliefs about illness, episodes in this category are heavily influenced by social and cultural factors. Physicians sometimes divide episodes of this kind into "necessary" and "unnecessary" visits. In doing so, the physician is assuming in the patient a knowledge of illness equivalent to his own. A patient who has for the first time felt his own xiphisternum and comes complaining of a lump is not displaying inappropriate anxiety. He has heard that any lump should be reported to the doctor. Only if he fails to respond to reassurance can his behavior be considered inappropriate. In this case, he should be placed in the third category since his symptoms probably conceal a problem of living. Of course, a patient may feel anxious about symptoms that have become intolerable. A myocardial infarct will drive a patient to his doctor and also may make him anxious. The patient is classified according to the predominant pattern of behavior. At a later stage, when symptoms are relieved and anxiety predominates, the patient may be moved from the first to the second category.

*Problems of living presenting as symptoms (Heterothetic)*—Many episodes of illness turn out to have their origins in some disturbance of the patient's relation with his environment. For example, most people experience some symptoms. Only a minority will be presented to physicians; the remainder will be ignored or self-treated. Whether or not a physician is consulted depends on several factors other than the severity of the symptom. In cases like these the question "What is the diagnosis?" is often irrelevant. Instead, the physician must ask questions like "Why did the patient come with this minor symptom?" Or "Why does this mother keep bringing her child with minor infections?" Relapses or new attendances for chronic disease may be due to a deterioration of the pathological process, or they may be due to a problem of living. Poor control of an adolescent's diabetes may be due to an identity crisis. A woman may come with her chronic backache because of the stress of her mother's illness. A woman may be seeking an operation for varicose veins, which she has had for 20 years, to win back the affections of her husband. All these episodes have two things in common: The patient presents with symptoms that conceal an underlying problem of living; and the clinical diagnosis is of secondary importance to the definition and understanding of the patient's problem of living.

*Administrative*—This category covers doctor-patient contacts whose purpose is administrative, even though the patient is ill, e.g., the provision of a certificate for an illness that would not otherwise lead to a demand for service.

*No illness*—This category includes all attendances for preventive purposes, such as antenatal or well-baby care, or for medical assessment when no symptoms are offered.

At any contact between patient and physician it should be possible to place the patient's behavior into one of these categories. At further contacts for the same illness, the classifi-

cation will remain the same unless the behavior changes.

### Patient Behavior and Psychiatric Morbidity

Many of the episodes are embraced by the term "psychiatric disease." Whether or not the disease model is appropriate to episodes of this type is a matter of debate. Some psychiatrists and social scientists argue that the decision to label a problem as an illness is social and cultural rather than scientific. By labeling a problem as an illness, society finds it easier to take a humanitarian rather than a punitive approach. We have witnessed this process with attempted suicide, with certain forms of juvenile delinquency, and, more recently, with unwanted pregnancy.

Whether or not this practice is defensible on scientific grounds need not concern us here. To me, it oversimplifies the problem in the mind of the physician. To turn a problem of living into an illness encourages us to think of it as a "thing" that has to be treated or removed, rather than as a complex breakdown of equilibrium between an organism and its environment.

Certain psychiatric syndromes are, of course, related to pathological processes in the nervous system. Others may be judged to have a physical basis that has not yet been demonstrated. For our purposes, these syndromes may be treated in the same way as other organic diseases. I believe that the vast majority of emotional disorders in general medical practice fall into the category of "problems of living" or of the anxiety that is natural to people who feel their health and well-being to be threatened. These phenomena are more appropriately classified as patient behavior than as psychiatric morbidity.

### A Taxonomy of Social Factors in Illness and in Patient Behavior

I have attempted to reduce to some sort of order a wide range of human responses to illness. A patient's behavior is determined by many factors, including his genetic makeup, his early imprinting, his previous experience of illness, his current life situation, and his aspiration for the future. Of all these factors, the current life situation is the most amenable to alteration by the physician, nurse, or social worker. My second proposal is for a classification of the interactions between the individual and his environment. We possess a vocabulary to describe some fairly static social factors, such as race and social class. But we have no way of classifying the ways in which a changing environment acts on an individual as he goes on his journey through life.

The relation of a sick person to his environment is complex. Even though not causal in the disease, social factors may determine the patient's behavior. A patient's illness and disability will itself produce changes in his life situation which may react on the patient to affect his behavior. We are not dealing here with linear relations but with a system. The following is an attempt to classify these dynamic relations between people and their human and material environments. Since a person may be acted upon by several processes simultaneously, the categories are not mutually exclusive. These factors can be termed "primary," if they are a cause of the illness or patient behavior, and "secondary," if they are a result of the illness.

*Loss:* Personal loss—loss of a loved one through death or desertion. Loss of things—imposed loss of home, cherished possession, or job.

*Conflict:* Interpersonal—conflict

within family, with neighbors, or at work, where hostility is recognized. Intrapersonal—role conflict or conflicting demands on the patient (as a working mother).

*Change:* Development—where time of life is the major problem (as in adolescence, menopause, or senescence). Geographic—where a move to an unfamiliar environment is the major problem (as an immigration).

*Maladjustment:* Interpersonal—problems between people with no overt conflict (as in failure to achieve a satisfactory sexual relation without hostility between partners). Personal—failure to adjust to the environment (home or job) in the absence of the above mentioned loss, conflict, or change.

*Stress:* Acute—unexpected event not covered under loss, conflict, or change (for example, the sudden illness of a family member or friend). Chronic—long-term situation not included in loss, conflict, or change (for instance, the presence of a handicapped child in the family).

*Isolation:* not due to any recent loss, change, or conflict (as in an elderly widow).

*Failure or frustrated expectations:* when the patient's goals in life are not fulfilled and when there is no evidence of an intervening event covered by loss, conflict, or change (e.g., failure at school or failure to achieve occupational promotion).

How can these taxonomic systems affect the way we think about illness? First, they provide us with a frame of reference and vocabulary with which to express our observations about patient behavior and the social factors in illness. Secondly, they release us from bondage of having to classify behavioral phenomena under the heading of diagnosis. Thirdly, they provide a system for classifying the many patient contacts in which there is no diagnosis in the accepted sense. Fourthly, they force us to think about the behavioral and social aspects of illness, just as the diagnosis forces us to think about the pathogenesis of disease. If physicians were expected to state the behavioral and social classification of every illness, as they are now expected to state the diagnosis, they would find it much easier to apply the behavioral sciences to their clinical practice. I believe that this new approach will enable us to study the epidemiology of medical practice in a way that provides a more accurate record of the facts. The next era in medicine may see much of the physician's role as a diagnostician taken over by the computer. If this occurs, the capacity of physicians to identify behavioral and social factors in illness may be one of their most important skills.

---

Condensed with permission from the *New England Journal of Medicine,* 287(8): 384–387, 1972. Copyright 1972 by the Massachusetts Medical Society. Reprints of original article from Ian Renwick McWhinney, M.D., Department of Family Medicine, Room 114, Medical Services Building, University of Western Ontario, London 72, Ontario, Canada.

# Validity of an Empirically Derived Phenomenological Typology

**John E. Overall, B. W. Henry, and James R. Markett**

Some method of classifying psychiatric patients is essential for organizing data and generalizing experience. An effective classification is one in which individuals who are placed in the same category are similar enough to be treated alike for most practical purposes. Operational definition and reliable criteria for assigning individuals among classification categories are the prime requisites of a pragmatic classification procedure.

Empirical methods, such as profile cluster analysis and homogeneous-grouping techniques, afford the possibility of developing useful classification schemes. The empirical methods can define subgroups of patients that are maximally homogeneous on the symptom characteristics subjected to analysis. Such methods can answer questions concerning *how many* and *what kinds* of homogeneous subgroups are present within a heterogeneous patient population. The important questions that remain concern the practical and scientific status of such empirically derived symptom clusters. Are patients who are classified in such manner similar enough to be treated alike for practical purposes? Can the empirical methods lead to identification of primary syndromes of psychopathology, or do they have only pragmatic value?

The psychiatric-medical concept of diagnosis includes more than phenomenological clustering of symptoms but involves an evolution of the concept of disease in which symptoms are linked to physiological processes. In view of the tenuous definition of most psychiatric diagnoses relative to etiology, physiological basis, predicable course, and prognosis, it becomes relevant to question whether empirically derived symptom clustering may not provide a more valid basis for definition of primary syndromes of psychopathology.

The primary advantages of an empirically defined phenomenological typology reside in operational definition and specified criteria that can result in more reliable assignment of individuals. The shortcomings of clinical diagnosis as a basis for classifying patients for research purposes have so frequently been noted that we will not recount them except to say that diagnostic criteria tend not to be clearly specified and diagnostic practices vary widely. Clinical diagnostic groups such as "paranoid schizophrenics" and "psychotic depressions" are phenomenologically heterogeneous, and there are indications that the phenomenological differences within diagnostic groups are important for clinical decisions such as treatment selection. Rather than belaboring the deficiencies of diagnostic practices, we ask the reader to recall his skepticism when he encounters reports based on diagnostic classification in which the results do not agree with his expectations. The first inclination is to question whether or not the sample was composed of true "psychotic depressions" or true "paranoid schizophrenics." This leads to an impossible situation as far as research or even communication of clinical experience is concerned. As an alternative, we propose a phenomenologically homogeneous and reliably distinguishable grouping of patients based on objectively specified criteria.

Then, it is important to investigate the pragmatic and scientific status of the empirically derived typology, as we have begun to do in this research.

The first author reported the results of a series of empirical profile-clustering analyses that lead to the identification of five distinct symptom patterns within a general psychiatric population. From the five empirically derived clusters, a phenomenological typology consisting of thinking disturbance, hostile extrapunitiveness, retarded depression, neuroticism, and agitation-excitement syndromes was developed. Each syndrome was described in detail by a Factor Construct Rating Scale (FCRS) profile, and a procedure for assigning individual patients among the five classes on the basis of FCRS profiles was provided. In the present investigation, the procedure for assignment of patients among the five phenomenological subtypes was employed with a new and larger sample of patients better to define the classification prototypes and to examine the validity of the phenomenological typology with reference to relevant external criteria.

Certain demographic and sociocultural characteristics have been demonstrated to be relevant with respect to the nature of manifest psychopathology. To the extent that a classification of psychiatric patients is adequate for characterizing important individual differences, it should produce subtypes that relate differentially to relevant background characteristics. Of greater importance in assessing the validity of a classification of psychiatric patients is the question of treatment assignment. Current diagnostic nomenclature has been criticized for lack of relevance in treatment decisionmaking. An effective and adequate classification scheme should be relevant for selection of appropriate forms of treatment. Finally, an adequate classification scheme should be relevant with respect to treatment outcome and subsequent prognosis.

A sample of 1032 new admissions to a psychiatric service over a period of 1½ years were subjects. The resident completed a Factor Construct Rating Scale (FCRS) within one week after admission on the basis of information obtained in routine clinical contact with the patients. The FCRS is shown in Figure 1. A senior medical student was employed to abstract from the clinical charts demographic data, treatment record, and final clinical diagnosis. Within the first week after hospital discharge, the resident responsible for the patient completed a final outcome evaluation form consisting of seven clinical ratings of improvement, outcome status, and subsequent prognosis.

Each patient was assigned to one of five phenomenological, symptom-descriptive categories on the basis of normalized vector-product similarity to classification prototypes. The method involves calculation of a coefficient of similarity (normalized vector product) between the patient profile and the prototype profile for each phenomenological class, with the patient being assigned to the class for which the similarity coefficient is largest.

Two-way contingency tables were constructed to examine the relationship of phenomenological classification to 15 background factors, treatment assignment, and to outcome and prognosis ratings. The background variables were selected on the basis of previous empirical investigations of their relevance for patterns of psychiatric symptomatology recorded on other rating instruments. The treatment variables in-

Figure 1. Factor Construct Rating Scale

Name _____

Rater _____

Date _____

Age_____ Sex _____

Education (Yrs.) _____

Marital Status (Divorced, Widowed, Second Marriage, etc.) _____

| | Extremely Severe | Severe | Mod. Severe | Moderate | Mild | Very Mild | Not Present |
|---|---|---|---|---|---|---|---|
| MENTAL DISORGANISATION—Confusion, disorganisation, impaired association or recall, disorientation. | | | | | | | |
| MENTAL DISTORTION—Bizarre distortions in perception or thought, hallucinations or delusions. | | | | | | | |
| AFFECTIVE IMPOVERISHMENT—Absence or diminution in emotional responsiveness, blunting. | | | | | | | |
| BIZARRE MOTOR MANIFESTATIONS—Odd, ritualistic or repetitive motor behaviour (exclude tics). | | | | | | | |
| PSYCHOMOTOR RETARDATION—Slow, deliberate, laboured or weakened speech and movements. | | | | | | | |
| AGITATION-EXCITEMENT—Excessive or exaggerated spontaneous motor activity, loud or hurried speech. | | | | | | | |
| EXTRAPUNITIVENESS—Vengeful intent, hostility, projection of blame. | | | | | | | |
| INTROPUNITIVENESS—Introjection of blame, guilt, self-criticism | | | | | | | |
| DEPRESSIVE MOOD—Despondency, sadness, pessimism, hopelessness. | | | | | | | |
| ELEVATED MOOD—Exaggerated high spirits, optimism, euphoria. | | | | | | | |
| PASSIVE DEPENDENCE—Seeking direction or reassurance, lacking in confidence and self-reliance. | | | | | | | |
| DOMINANT CONTROLLING—Aggressively confident, directing, domineering. | | | | | | | |
| DENIAL—Unwillingness to admit or inability to perceive problems, conflicts or psychopathology. | | | | | | | |
| INTELLECTUALISATION—Superficial explanation without true understanding or emotional acceptance. | | | | | | | |
| OVER-RESPONSIVENESS—Dramatised, excessive, superficial emotional responsiveness. | | | | | | | |
| NEUROTICISM—Obsessive, compulsive, or phobic neurotic symptoms. | | | | | | | |
| ANXIETY-TENSION—Uneasiness, worry, fear or overconcern usually with nervousness and tension. | | | | | | | |

* The spelling of certain rating constructs reflects original publication in a European-based Journal where the "King's English" is preferred by editors.

Table 1. Mean FCRS Profiles for Five Phenomenological Sub-Types

| | I | II | III | IV | V |
|---|---|---|---|---|---|
| Mental disorganization | 0.6 | 3.5 | 0.8 | 0.6 | 1.4 |
| Mental distortion | 0.4 | 3.5 | 1.4 | 0.5 | 1.3 |
| Affective impoverishment | 2.1 | 2.2 | 1.3 | 0.6 | 0.3 |
| Bizarre motor manifestations | 0.2 | 1.2 | 0.2 | 0.3 | 0.6 |
| Psychomotor retardation | 2.0 | 1.3 | 0.4 | 0.4 | 0.1 |
| Agitation-excitement | 0.3 | 1.3 | 0.6 | 1.4 | 3.0 |
| Extrapunitiveness | 0.4 | 0.7 | 2.0 | 1.0 | 1.6 |
| Intropunitiveness | 1.5 | 0.8 | 0.5 | 1.5 | 0.6 |
| Depressive mood | 3.2 | 1.3 | 1.2 | 2.2 | 1.0 |
| Elevated mood | 0.0 | 0.1 | 0.0 | 0.1 | 2.1 |
| Passive dependence | 1.4 | 0.8 | 0.9 | 2.6 | 0.4 |
| Dominant controlling | 0.1 | 0.2 | 0.6 | 0.3 | 1.8 |
| Denial | 0.7 | 0.9 | 2.2 | 1.4 | 1.7 |
| Intellectualization | 0.3 | 0.4 | 1.5 | 1.2 | 0.8 |
| Over-responsiveness | 0.2 | 0.3 | 0.4 | 1.5 | 1.1 |
| Neuroticism | 0.4 | 0.2 | 0.5 | 2.2 | 0.4 |
| Anxiety-tension | 1.9 | 1.4 | 1.7 | 2.7 | 1.5 |

cluded type of drug, use of ECT, and psychotherapy. The outcome and prognosis ratings were those obtained from the psychiatry residents at the time the patients were discharged.

The mean FCRS profiles for patients assigned to each of the five phenomenological profile types are presented in Table 1. To the extent that these mean profiles differ from the original empirically derived prototypes that were used on the basis for the classification, the new mean profiles can be considered more representative of the natural configuration of patients in this psychiatric population. This conclusion is offered not only because of the larger sample involved, but also because the new means have the least average distance from all patients who were assigned to each category. In this respect, the new mean profiles may provide a superior basis for phenomenological classification based on FCRS profiles.

Although we do not consider clinical diagnosis to represent a uniquely valuable criterion against which to judge the validity of classification results, it is worthy of note that the empirically derived profiles correspond quite closely to conceptual stereotypes of major diagnostic classes. If one considers schizophrenia, depression, mania, neuroses, and character disorders to represent the major subdivisions of functional psychiatric disorders, it is impressive that the empirical clustering methodology resulted in one profile pattern suggestive of each major diagnostic division. This observation is taken as one bit of evidence supporting the validity of the clustering methodology for identification of primary syndromes of psychopathology.

The correspondence between final clinical diagnosis of patients in this study and their phenomenological profile classification was examined by cross-tabulation in a two-way contingency table. Clinical diagnoses were grouped into schizophrenias, depressions, neuroses including psychophysiological reactions and various personality disorders, excepting psychopathic and sociopathic personality disorders which were placed in a separate character disorders category. Patients with organic brain syndrome and other organic diagnoses were also placed in a separate category. Patients with depressive diagnoses were predominantly depression profile types, although a substantial number had neuroticism profile patterns. The high incidence of neuroticism profile patterns among the depressive diagnostic groups is not surprising, particularly since "neurotic depression" was included in the general diagnostic category. Patients with schizophrenic diagnoses were

most often classified phenomenologically into the thinking disturbance category, although substantial numbers were placed in each of the other phenomenological categories. Patients with psychoneurotic and related diagnoses were most often classified phenomenologically into the neuroticism category, but substantial numbers had profiles which fell into the extrapunitiveness and depressive phenomenological classes. The psychopathic character disorder diagnoses were most often extrapunitiveness profile type, and the patients with manic diagnoses were almost all of the agitation-excitement profile type. The patients who carried organic diagnoses were predominantly of the thinking disturbance phenomenological type. The correspondence between diagnosis and phenomenological classification is interesting, and appears to add evidence that empirically derived clusters may represent primary syndromes of psychopathology. The lack of a stronger relationship to clinical diagnosis does not surprise us in view of what we know about diagnostic practices. It does emphasize, however, that phenomenological classification and clinical diagnosis are two different things.

The correspondence of the phenomenological typology represented by the FCRS profiles to typologies derived empirically from analyses of other kinds of symptom ratings provides additional evidence of the utility of cluster analysis methods for identification of primary syndromes of psychopathology. In analyses involving clustering of Brief Psychiatric Rating Scale Profiles (BPRS), six phenomenological types were identified. They were described as florid thinking disorder, withdrawn-disorganized thinking disturbance, paranoid hostile suspiciousness, anxious depression, hostile depression, and retarded depression. In analyses of data from a study of the chemotherapy of mania, an agitation-excitement pattern was added to the BPRS typology. In the FCRS, the detail of symptom description is not adequate to distinguish florid psychotic thinking distortion from the disorganization (syntax) disturbance; thus a single major thinking disturbance syndrome evolved from empirical clustering of FCRS profiles. The FCRS extrapunitiveness patterns seem to correspond to the hostile-suspiciousness syndrome represented in BPRS profiles. The depression subtypes defined from analyses of FCRS profiles correspond to the retarded depression pattern derived from analyses of the BPRS. What we call "neuroticism" appears to correspond to the simple anxious depression cluster that has been observed in analyses of BPRS profiles. Finally, the agitation-excitement pattern corresponds to a similar pattern defined for the BPRS. The fact that the same general phenomenological patterns tend to result from analyses of different symptom rating profiles lends credence to the notion that they may represent basic syndromes independent of the rating scales from which they are derived.

### Background Characteristics

The distributions of background characteristics within the five phenomenological subtype categories were examined by constructing two-way contingency tables. The following statements convey the primary findings:

*Age*—Relatively high incidence of depressive profile patterns in the 30–50 age range. Highest incidence of thinking disturbance after age 60 and next in youngest age group. Extrapunitiveness pattern most prevalent in

younger groups. Incidence of neuroticism patterns highest in 20–40 age range. Agitation-excitement highest in 40–50 decade, with relatively high incidence thereafter.

*Ethnic minority*—Thinking disturbance profile patterns relatively more frequent in Negro and Latin minority ethnic groups. Depressive, agitation-excitement, and hostile extrapunitive patterns more frequent in Anglo majority group.

*Sex*—Depressive, thinking disturbance, and neurotic patterns more frequent in females. Hostile extrapunitiveness and agitation-excitement more frequent in males.

*Age onset*—Onset of illness for patients with depressive profile patterns most frequent in middle years. Thinking disturbance onset relatively frequent before age 20 or after 50. Onset for neuroticism patterns most frequent in 20–40 age range, and extrapunitiveness pattern onset most frequent before age 20.

*Duration present episode*—Longer duration of present episode more common in depressive syndromes and shorter current episode more common in thinking disturbance syndromes.

*Previous hospitalization*—Previous psychiatric hospitalizations least common among patients with depressive profile patterns, most common among neurotic and agitation-excitement groups, with hostile extrapunitiveness and thinking disturbance group in between.

*Education-graduation*—Completion of college or professional school more frequent among patients with agitation-excitement profile patterns and next most common in neurotic group. Completion of college least common among patients with thinking disturbance profile patterns.

*Marital status*—Single (never married) status more common among thinking disturbance and extrapunitiveness profile groups, least common among agitation-excitement and depressive profile types.

*Divorce history*—History of divorce most common in neurotic profile types, and least common in the thinking disturbance profile types who do become married.

*Children*—Relatively larger proportion of agitation-excitement and depressive profile types have children, with two-four children most common among depressive types and one child relatively more common among agitated types.

*Religious involvement*—Excessive or fanatic religious involvement most frequently noted among neurotic profile types, with lowest frequency of excessive religious involvement among depressive and agitated types.

These results are not presented to document unknown relationships of psychopathology to background variables. The particular background variables were chosen because of their familiar relevance for psychopathology. The fact that the phenomenological typology relates to the background variables in an unexpected and meaningful way is offered as indication of the validity of the empirically derived typology. Patients who are grouped together tend to have similar demographic characteristics and similar social and psychiatric histories. Age of onset, duration, and course of illness tend to be systematically different for patients in the different phenomenological categories. Such relationships provide an essential, if not sufficient, basis for considering the phenomenological types to represent major syndromes of psychopathology.

### Treatment Decisions

The next aspect of clinical relevance involved treatment decisions.

A contingency table was constructed to evaluate the relationship between phenomenological profile classification and treatment assignment. Choice of treatment was highly related to the phenomenological classification of the patients being treated.

There was a consistent tendency to use tricyclic antidepressant drugs in the treatment of patients whose symptom rating profiles were classified as depressive type. Tricyclic antidepressants were also used with relatively high frequency in the treatment of patients classified as having neuroticism profile patterns, and they were used quite infrequently in patients with thinking disturbance or extrapunitiveness profile patterns. The major tranquilizers were used with greatest relative frequency for patients with thinking disturbance and extrapunitiveness profile patterns, and also in the agitation-excitement group. Proportionately, patients with depressive and neuroticism profile patterns were treated less often with major tranquilizers. Minor tranquilizers were used to treat patients with neuroticism and depression profile patterns. Finally, of the patients who were treated with lithium carbonate, a substantial majority were classified as having agitation-excitement profile patterns. Thus, the drug usage for this patient population strongly supports the validity of the profile classification scheme and the interpretations of the syndromes that the five profile patterns represent.

ECT was employed most often in the treatment of patients with thinking disturbance profile patterns and next most often in the treatment of patients with depressant profile patterns. ECT was used least often in the agitation-excitement group, with only slightly greater relative frequency in the extrapunitiveness and neurotic profile types. The frequent use of ECT in treatment of patients with thinking disturbance syndromes has been reported in surveys of practices in our inpatient treatment setting; thus the observation should not be interpreted as casting suspicion on the classification grouping. In this hospital setting, ECT is used to treat severe thinking disorders and depressive syndromes, and the simple profile categorization appears adequate in distinguishing groups with high and low probability of receiving ECT.

Psychotherapy was employed to treat a majority of patients whose profiles were classified as neuroticism type. Next in frequency were patients whose profiles were classed as extrapunitiveness patterns, and next those with depressive patterns. Psychotherapy was employed with lowest relative frequency in the treatment of patients whose profiles fell into the thinking disturbance category. These results appear consistent with the interpretations placed on the empirically derived profile patterns. The simple phenomenological classification appears adequate to group patients according to probability of receiving psychotherapy, as well as drug and ECT treatments.

**Prognosis**

The final series of analyses was undertaken to evaluate the relationship of clinical ratings on seven outcome and prognosis variables to the phenomenological classification. Some of the clinical ratings concern immediate discharge status while others involved prediction of future adjustment. Disregarding questions concerning the validity of the clinical ratings, it should be expected that a valid classification scheme should result in a grouping of patients which is at least modestly related to subsequent outcome and prognosis evaluations. It

should be remembered that the symptom ratings on which profile classifications were based were obtained within the first week following admission. The discharge status and prognosis ratings were obtained after the patient's discharge from the hospital, usually several weeks later.

The profile classification appeared substantially related to subsequent discharge evaluation and prognostic estimates, with exception of the global improvement rating. It should be noted that the global improvement rating was made with reference to patients of "the same diagnostic type," which implies an effort to remove diagnostic differences from the consideration of response. Thus, it may not be surprising that the profile classification did not relate significantly to that variable.

Two of the rating questions involve similar judgments concerning the need for continued support and care. We will characterize these as self-reliance ratings. On these scales, patients with thinking disturbance profile patterns received substantially poorer ratings than other groups. The proportion of very good self-reliance ratings was greatest for patients with agitation-excitement profile patterns, and next greatest for neuroticism and extrapunitiveness profile types. The relative frequency of intermediate to good self-reliance ratings was highest for the depressive profile types.

The relationship of residual pathology to profile type was a weak one, presumably because judgments were made with reference to patients of the same diagnostic type. Predictions of "future adjustment" were most favorable for the agitation-excitement type, and least favorable for the thinking disturbance type. In the question of "approach to maximum capacity," again the agitation-excitement type appeared best, with the hostile extrapunitive type next. Without going into detail, we can summarize by saying that the phenomenological descriptive classification based on symptom rating profiles obtained soon after admission appeared relevant with respect to the clinical judgments made several weeks later when the patients were discharged from the hospital. Such predictive validity supports the pragmatic utility of the classification scheme, and it lends another bit of support to the notion that these empirically derived classes may indeed correspond to primary syndromes of psychopathology in a more fundamental manner.

---

Condensed with permission from the *Journal of Psychiatric Research*, 9(2): 87–99, 1972. Copyright 1972 by Pergamon Press, Ltd. Reprints of original article from John E. Overall, Ph.D., Department of Neurology and Psychiatry, The University of Texas Medical Branch, Galveston, Tex. 77550

# Syndromes of Psychosis and Nonpsychosis: Factor Analysis of a Systems Analysis

**Raymond D. Martorano and Peter E. Nathan**

Dissatisfaction with the results of clinical decisionmaking continues to preoccupy many mental health professionals. Although responsibility for inadequate clinical decisionmaking has been apportioned to the clinician and to the diagnostic instruments and other evaluative procedures he uses, utility of traditional Kraepelinian diagnostic groupings continues to be debated. One result is a recent overhaul of the "official" psychiatric classification system in an effort to make it more useful for predictive purposes. Another indication of dissatisfaction with the "traditional" labeling system is an increase in efforts to evolve empirically based diagnostic systems.

In an effort to identify the most common sources of diagnostic unreliability, the second author of this paper undertook a systems analysis of traditional diagnostic procedures. This effort identified a set of 100 psychiatric signs and symptoms of key differential diagnostic significance. The systems analysis also traced the diverse pathways by which these cues are processed by clinicians for diagnostic purposes. The Boston City Hospital-Behavior Checklist (BCH-BCL), a 100-item symptom checklist, is a direct translation of the systems analysis into a form designed to enable investigation of the validity of that analysis.

This paper reports the factor analysis of symptom data from 924 psychiatric patients whose psychopathology was rated on the BCH-BCL. Beyond its value as an exploration of the dimensions of psychopathology which underlie the major psychological disorders, this factor analysis is of interest because the BCH-BCL includes a number of items specifically tapping nonpsychotic behavior. Hence, symptom data from this instrument might define syndromes ranging through the entire range of psychopathology. Since the data were obtained from inpatients and outpatients, whose psychological disabilities ranged from very severe to very mild, the likelihood is increased that syndromes defined might describe general parameters of psychopathology rather than those only of psychosis.

Each of the 924 psychiatric patients in the sample was evaluated by one of 16 experienced psychiatric residents at the Boston City Hospital between late 1966 and early 1968. Patients ranged in age from 16 to 82 (mean age, 34). Approximately 30 percent were black or Puerto Rican; almost 40 percent were males.

Patients were evaluated within 24 hours of self-referral to the hospital accident floor or within 24 hours of admission to its psychiatric inpatient ward. Each patient was given a complete psychiatric evaluation. Each examiner completed a BCH-BCL at this time. Every resident was asked to consult a senior staff person when he was unsure of a diagnosis or a plan for disposition; such consultations were always available in practice and were often employed in fact.

The BCH-BCL includes six separate sections, each exploring a relatively homogeneous group of symptoms. The first section looks for symptoms thought to differentiate among the board diagnostic categories (e.g., psychosis, psychoneurosis); the succeeding five sections record

the presence of symptoms of abnormal psychomotor, perceptual, cognitive, and affective behavior, and disordered levels of consciousness. These symptoms are commonly thought to enable decisions on specific diagnostic labels (e.g., paranoid schizophrenia, reactive depressive). BCH-BCL items are dichotomous: They ask only whether or not a symptom characterizes the behavior of the patient being evaluated. Dichotomous items call for a positive item decision only when the examiner has actually observed the behavior defined by the item. Despite this provision and the fact that all raters were trained to rate only behavior they actually observed, one cannot state unequivocally that no contamination of diagnostic statements resulted from information on patients' social history.

After all data had been keypunched and verified, a contingency analysis of all items was performed. As a result, 30 of the original 100 items were withdrawn from further analysis, either because they appeared with very low frequency in the sample population or because they intercorrelated highly with another item which described essentially the same symptom.

Correlations among the remaining 70 items were computed and the matrix of intercorrelations was factored. Two rotations, both based on the oblimin rotation criterion, were performed. In the first instance, an orthogonal, varimax rotation was accomplished. In the second, the oblimin criterion was applied to the primary factor loadings to allow the factors to be oblique and thus yield a simple loadings rotation. Only the results of the oblique rotation are reported and discussed.

The factors recovered by the oblique rotation are as follows: Factor 1 loads heavily on symptoms reflecting the acute cognitive and affective disorganization most characteristic of Schizophrenic Disorganization. Factor 2 taps a wide range of symptoms of Depression. Factor 3 reflects Disordered Memory and Consciousness, a complex of symptoms which often cohere in acute brain disorder. The Mania-Hypomania syndrome is described by Factor 4. Factor 5 contains most of the cardinal symptoms of Catatonia, including catalepsy, perseveration, and negativism. Factor 6 taps subjective Anxiety. Factor 7 includes psychotic symptoms describing Abnormal Perceptual Behavior. Factor 8 defines a syndrome most appropriately labeled "Exogenous" or "Reactive" Depression, while Factor 10 circumscribes the dichotomous syndrome, "Endogenous" Depression. Factor 9 loads on Autism associated with Stupor, a syndrome often associated with toxic conditions. Factor 11 includes symptoms of Abnormal Affective Behavior, a syndrome most often associated with severe affective psychosis, while Factor 12 seems to tape many of the central elements of the Paranoid Process. Factor 13 reflects some of the symptoms of Abnormal Cognitive Behavior, symptoms which are often observed in the acute organic disorders.

Fifty-seven percent of patients given the diagnosis of paranoid schizophrenia achieved high loadings on Factors 7 (Abnormal Perceptual Behavior) and 12 (Paranoid Process); 36 percent, 27 percent, and 35 percent of patients in the same diagnostic category loaded on Factors 1 (Schizophrenic Disorganization), 4 (Mania-Hypomania), and 11 (Abnormal Affective Behavior). By contrast, 45 percent of Anxiety Reaction patients loaded highly on Factor 6 (Anxiety), the only factor on which more than

25 percent of patients with this diagnosis achieved significantly high or low scores. Relatively few patients given any of the Personality Disorder diagnoses showed high factor loadings. Patients with antisocial reaction, passive-aggressive personality, and emotionally unstable personality diagnoses failed to load highly on any factor, while moderate numbers of patients in all but two other groups within this diagnostic category loaded on but one factor, that circumscribing "Reactive" Depression.

**The Validity of "Traditional" Syndromes**

Examination of the matrix of intercorrelations among factors permits observations on the validity of "traditional" diagnostic groupings. Factors 1, 4, 7, and 9, all highly intercorrelated, circumscribed many major symptoms of the schizophrenias; Factors 3, 4, 7, 9, and 13, also highly related, describe behavior characteristic of patients given Acute Brain Syndrome diagnoses. Given these data, one might speculate that the two conditions, extremely difficult to differentiate from one another in the clinical setting, are so in part because they share three major syndromes (Factors 4, 7, and 9). One might also conclude that symptoms of Schizophrenic Disorganization (Factor 1) and Disordered Consciousness and Memory (Factor 3) might be most useful in discriminating between schizophrenic and toxic conditions because the two diagnostic groups do not share these two syndromes.

The utility of current diagnostic practices vis-a-vis the two major neurotic disorders, Anxiety Reaction and Depressive Reaction, was just partially supported: Only 45 percent of Anxiety Reaction patients attained elevated scores on Factor 6 (Anxiety) and only 33 percent of Depressive Reaction patients scored high on Factor 8 ("Reactive" Depression). While these factors were the only ones which sizable numbers of neurotic patients loaded highly, the fact that relatively few of these patients loaded highly on them was surprising, since we had assumed that the symptoms included in these two factors were virtually pathognomonic of the two neuroses. On the other hand, these unexpected findings highlight the difficulty clinicians often have differentiating these conditions from others with prominent depressive or tension-filled features. In this connection, it is interesting that four other diagnostic groups demonstrated high factor loadings on Factor 6, while five others did so on Factor 8. Most of these groups were Personality Disorders, that set of conditions with which the neuroses are most often contrasted during diagnostic decisionmaking.

**Limitations on Generalizability**

Certain design and sample characteristics restrict the extrapolation of results to other populations. Most of our sample was poor, underemployed, and undereducated. Furthermore, the raters made their diagnostic decisions on the basis of all the information at their disposal. This included, but was likely not limited to, data tapped by the BCH-BCL. Finally, it is possible that another factor solution could have been obtained from our data which might have better explained and organized them.

---

Condensed with permission from the *Journal of Abnormal Psychology*, 80(1): 1–9, 1972. Copyright 1972 by the American Psychological Association, Inc. Reprints of original article from Peter E. Nathan, Ph.D., Psychological Clinic, Rutgers University, New Brunswick, N.J. 08903

# Studies of Clinical Judgment: Part II. Toward A New Model for Research

**Michael K. Harty**

I previously reviewed approaches to the study of clinical judgment and found them inadequate to deal with many judgments confronting clinicians; that is, judgments concerning inner experience. It is my thesis that research methodology should encompass judgments in which the humanness of the judge may play an important role—the judgments about inner experience. My aim here is to develop a model of the judgment process incorporating this idea, and then to advance some methodological proposals based on it.

First, I will indicate how "we are all pretty much alike" for purposes of psychological assessment. Next, I will draw support from research on empathy for the existence of "emotional knowing" as a measurable phenomenon. Finally, I will describe findings from research on person perception which are relevant to the framework of ego-psychology. Then, I will propose a methodology for further research.

### Universality of Affective Experience

For the statement, "We are all pretty much alike," to serve as a basis for a theory of understanding the experience of another, it must mean something like: There is a set of affective experiences common to every human being. These may be consequences of universal basic needs, of human physiology, and/or of universal kinds of dealings with the environment, but they are independent of particular learning experiences and must be regarded in some sense as universally present. These experiences affect behavior in a lawful fashion so that all behavior, to a greater or lesser extent, constitutes an expression of the affective state of the individual.

It is not necessary to go into detail about how the "basic emotions" might be classified, but an additional assumption is made about the manner in which they affect cognitive processes: Every variety of affective experience in the course of the individual's development becomes inseparable from images, fantasies, etc., having to do with the situations and the motivational context, conscious or unconscious, in which the affect has arisen; and these images and fantasies reflect the unconscious aspects of motivation according to the principles of "primary process," which are universal. Thus the relevant similarities between the judge (clinician) and the object (patient) lie in the potential availability to both of a common set of affective states and in the fact that affective states organize expressive behavior in similar ways in both of them. This is not to deny that the particular images, etc., which a person employs are largely determined by the cultural context and by the individual's history. The "vocabulary" involved in the expression of subjective states is highly variable, and the overlap between the subject's and judge's "vocabulary" will be an important factor in the judge's understanding. However, the "syntax" of this process is shared by every person. In the clinical situation, the clinician has available to him a sample of the patient's behavior, perhaps including fantasy productions. The clinician may add to his understanding of the patient's subjective state by temporar-

ily "stepping into the patient's shoes" and experiencing the patient's expressive behavior as if it were his own. The process by which the feeling state of another has been grasped has been referred to as empathy.

**Research on Empathy**

Psychoanalytic writers have described empathy as an "emotional knowing," which, in its mature form, involves a splitting of the clinician's ego into a participating and observing part and an oscillation between these two positions. It is akin to creative endeavor in that it involves "regression in the service of the ego." "Empathy" need not be treated as mysterious and unscientific simply because it refers to a complex hypothesized internal process, but may be systematically considered in studies of clinical judgment.

**Research on Person Perception**

Clinical judgments in relation to the personality and cognitive style of the clinician who makes them can be studied. It is necessary to look to research on "person-perception" for some relevant findings. "Person-perception" research and "clinical-judgment" research are not very different: Any unique contribution of the clinician must come from characteristics or abilities he possesses by virtue of being human. These characteristics or abilities would appear also in the general population from which the subjects of "person-perception" studies are drawn. Trained clinicians should be better able to use these abilities in forming judgments, but this is an empirical question. For present purposes, it will be sufficient to note some things which have been learned about the way people form their impressions of others:

- Forming impressions of others is an organized process, not simply a summing up of bits of information. Something like a structure is grasped; some cues are more central or influential than others, and the meaning and importance of particular traits is influenced by the configuration of other traits.
- There are stable individual differences in the way perceptions of others are organized. The principles according to which perceptions of others are organized represent the perceiver's "stereotype" or "implicit personality theory."
- One's "implicit personality theory" reflects more general cognitive tendencies. Chief among these tendencies is the "complexity-simplicity" dimension.
- Different kinds of cognitive organizations are effective for different tasks. Statements regarding the accuracy of interpersonal perceptions are difficult to make, since this area is plagued by the same criterion problems as is the study of clinical judgment. Nevertheless, an extensive review of the literature concluded that the type of subject, the relationship between judge and subject, the type of judgment, and the information available all enter into the likelihood that a particular judge will succeed, although there is also a general factor involved.
- Good judges of others are distinguished by certain personality traits. Good judges tend to be "better adjusted" than poor judges on whatever global measure is being employed.

**Proposed Model of Clinical Judgment**

The model is that the clinician progressively builds up a new substructure in his own ego which corresponds to the psychic world of the patient, and that the potentially unique con-

tribution of the clinician to psychological assessment derives from his ability to draw upon his own experience in such a way as to enrich, extend, and elaborate this model in the direction of relating more and more of the patient's observable behavior to underlying subjective states. His task involves something analogous to learning the patient's language—the personal idiom of images, wishes, fantasies, behaviors, etc., in which the patient expresses his version of human experience.

Whether such a contribution is needed will depend on 'economic" factors—the cost-payoff matrix associated with the task at hand. In most cases where a prediction of a specific overt behavior is needed, this kind of clinical judgment is not preferred, particularly when uncontrolled environmental factors are involved. Establishing the conditions for effective clinical judgment under such circumstances would be comparable to learning a whole language in order to translate a single word. For most purposes, it would be preferable to consult a dictionary and settle on the most common meaning—to operate in an acturial fashion. Similarly, where the aim is a diagnostic categorization, there is no necessity for learning the patient's language; what is called for is an application of the clinician's own language system to the patient's behavior as observed data. However, when the need is for something analogous to a phrase-book (say, a psychological report) to assist a traveler (therapist) in finding his way in a new territory (the patient's symptoms and other behavior) and in avoiding unnecessary clashes with the natives (therapeutic blunders), the learning of the language seems justified.

The process of building up an internal representation of the patient begins when the clinician registers some bit of the patient's behavior. To begin with, the clinician may have nothing but his own "language" to use as a first approximation of the set of relations between the patient's behavior and his inner experience. At this point, similarity between patient and clinician as to cultural background, social class, education, and other factors which influence expressive style is likely to facilitate the building of an accurate model, However, similarity between clinician and patient also invites inaccuracy by virtue of the clinician's assuming more similarity than actually exists. This is where the clinician's reality-testing must enter as a check on his empathy. He must be able to maintain the tentativeness of the model sufficiently to allow new observations to be used in two ways: as checks on the existing model and as potential aids in altering it.

As a wider range of patient behavior becomes available to him, the clinician has more opportunity to alter his model of the patient; however, he also has more opportunity for error, since his own "blind spots" have an increased chance of affecting his judgment. Thus the clinician who has a need to deny certain areas of his own experience may refuse to pick up cues which might be interpretable in terms of those areas of experience. Or he may pick up the cues but deny, minimize, exaggerate, alter, or distort their significance.

Although primary-process mechanisms are involved in the empathic process by which the internal model of the other is formed, the mature "generative" form of empathy results in "secondary-process object representations." This is a term for the more general case of "the clinician's internal model of the patient"; it is char-

acterized by relative accuracy, stability, a preserved temporal and spatial (external) index, and reflective processing.

Besides anxiety-laden concerns about uncertain ego boundaries, there are other factors which might interfere with the ability to have access to the internal representation of another. Among these are superego prohibitions against the necessary regression, strong needs to defend against the psychosexual implications of the act of observing another person, and strong needs to perceive the other in accordance with one's own motivations.

It is important that the suspension of reflective self-representations be reversible. The whole process of forming brief identification with the other is primarily a source of hypotheses. The hypotheses derived can then be checked against information available from other sources, phrased for inclusion in a test report, used as a premise in a deductive inference, or whatever the situation calls for. Implied here is the position that accurate perception of the other's experience is conceptually separate from the ability to communicate such perception.

**Application to Psychological Testing**

Everything said so far has been in terms of a generalized situation—interview, brief encounter, extended psychoanalysis, or whatever. A few words about psychological testing seem in order. The clinician enters the testing situation with one relatively stable "internal model": his internal representation of the test situation as it affects the patient. This may be largely in terms of the instrument he uses, the circumstances under which tests are administered, his own effect on patients, or any combination of these. Because the environment confronting the patient is being manipulated in a way familiar to the clinician, the clinician can juxtapose his existing model of the testing environment with his developing model of the patient's experience to gain some idea of what the patient is reacting to. For example, the clinician notices anxiety in the patient while administering the Rorschach; the clinician's knowledge that a certain card often evokes an aggressive image, along with the fact that the patient did not produce such an image, leads him to hypothesize that the patient became anxious about an aggressive fantasy which was denied conscious representation.

A second aspect of the psychological testing situation is that the patient's responses are preserved in permanent form. This makes it possible for the psychologist to check and recheck his model of the patient against the events which occured during the testing. This feature constitutes a potential advantage over the interview, where the clinician sees each bit of data once. It is a potential disadvantage, primarily the reduction of all responses to one modality and the limitations imposed on the patient's responses by the choice of tests. It is important for the clinician to remain aware that such limitations do exist—that test data do not reflect every aspect of the patient's experience.

There is no a priori reason why test scores should not qualify as "bits of behavior" on the basis of which inferences about experience might be made. It would seem, however, that the use of numerical scores in this way would be the exception, and would depend upon long experience with the instrument and the building up of a private system of experiential referents for various scores, The "empathic" approach seems more likely

to be useful with tests which are not used as "tests" but as "data elicitation techniques" which provide material for the interpretation of meaning. Scores in general belong more to the realm of logic, of after-the-fact checks on the plausibility of empathically gained hypotheses.

**Some Proposals for Research**

This section is in three parts: a reconsideration of the criterion problem; a discussion of alterations in methodology which would enable new light to be thrown on some of the "classical" questions about clinical judgment; and some research questions which grow out of this model along with suggestions for their empirical solution.

**The Criterion Problem**

In most research on clinical judgment, the experimental method has dictated the task to be studied. The conception of psychological science as requiring direct predictions of observable behavior has led researchers to gloss over the fact that "clinical judgment" can refer to a wide variety of tasks, and to act as though the tasks most easily studied (those calling for a concrete behavioral prediction) were representative of the whole field. Recognizing this, however, leaves us no better off than before, unless a criterion more satisfactory than directly observable behavior can be found. Research on judgment is severely handicapped if there is no way of distinguishing accurate judgments from inaccurate.

How can we establish a criterion of correctness of judgments about the inner experience of another person—which is directly observable only to the person himself, and which even the experiencing person is often unable to report accurately or coherently? Problems of this sort are not new; they have been handled in one of two ways. The first might be called "inferential": The consequences of correct and incorrect judgments are deduced and judgments are then evaluated on the basis of occurrence or nonoccurrence of these consequences. The more commonly used criterion is one of consensus among judges specifically trained for the purpose.

The approach suggested here is a combination of these two criteria. In the context of an "inexact science," where it is not possible to specify completely the evidence employed in forming a judgment or the "background information" relevant to the judgment, the best indication of the "true state of affairs" is often the judgment of an expert. Clinical judgments about the inner life of a patient fall in this category. "Expertise" is not simply a matter of reputation or arbitrary selection; rather, it can be established empirically on the basis of performance with the task in question. In some cases this would mean that the judgments of the "expert" would be subjected to the same kind of evaluation dismissed above as too cumbersome for routine use. The procedure would be equally cumbersome, but with the difference that it would need to be applied to only one person (or a few people). The time and effort involved in such an evaluation would be spent developing a measuring instrument, rather than simply making a measurement; once established empirically, the expert would become a standard in his own right.

Expert judgment could be used as an "objective estimate" of the probability values which ought to be attached to various possible states of affairs. This procedure could be applied to clinical judgments without

seriously altering the clinician's activity; many statements clinicians make about patients' experience have a quasi-probabilistic form. Once expert opinion has been collected in the form of probability statements, the accuracy of the judgments of the clinician-subjects could be assessed in at least two ways. If each clinician is required to make a single statement or description, accuracy could be assessed by determining what probability the expert attaches to the state of affairs described by the clinician.

A second approach would be to elicit from the expert a set of alternatives, with attached probabilities, and to use as the criterion of a clinician's correctness the degree to which the probabilities he assigns to the alternatives match those assigned by the expert. Neither of these approaches restricts the expert to the information available to the clinician-subjects; if the emphasis is on accuracy, and the expert is being used as an objective indicator of the "true state of affairs," it is important that he have available as much information as he needs.

### Some "Classical" Issues

Several topics have been studied in relation to clinical judgment: the relation of the clinician's accuracy to the amount of information he is given; the relation of accuracy to training and/or experience; the reliability of clinical judgments. Simply introducing a new criterion of accuracy which allows the study of judgments about others' experience is not the only methodological shift called for by this concept of clinical judgment.

The main point is that inner experience can be translated into action in an infinite number of ways; a method which requires the clinician to predict the subject's specific translations misses the point of much clinical judgment, which is to translate actions back into the inner experience which gives rise to them. Such an enterprise may require intensive, rather than extensive, information. For purposes of making inferences about a subject's inner state, it may be more useful to have a record of all his thoughts during a 5-minute period than to have a detailed account of his life history.

There is a related point with respect to the relationship between accuracy and clinical experience or training. The typical findings here have been as uncomplimentary to clinicians as the clinical-versus-statistical findings, and have in some cases come from the same studies. Once again, however, we must note that the choice of task seems to have been dictated more by the need for a nonarguable criterion than by a desire to illuminate the cognitive processes involved in successful clinical work. Investigators frequently maintain that their tasks are chosen to correspond with the clinician's habitual activity, but in most cases the resemblance seems superficial and more a matter of similar materials and terminology than of similar processes. What this points up is the need for more careful specification of the experience of the judges with the kind of task being performed. It seems likely that as tasks closer to the actual experience of clinicians are studied, the superiority of experienced clinicians will be demonstrated.

"Reliability" of clinical judgments can be defined in more than one way: stability, or reliability across time (for the same judge using the same data); consensus, or reliability across judges (for the same data and the same occasion); and convergence, or reliability across data sources (adminis-

tered on the same occasion and interpreted by the same judge).

How should reliability be defined in studying judgments about inner experience? The definition of reliability which has been least favorable to clinicians—convergence—is least appropriate for this type of judgment. It reflects a conception of personality as a collection of static traits which exist in some permanent sense and ought to be measurable by a variety of methods. The conception employed in making clinical judgments about inner experience is the antithesis of this; the interest is in when or under what conditions certain "traits" make their appearance, and each new "data source" represents a different set of conditions designed to help answer this question. Either of the other two definitions of reliability could be appropriate.

**Issues Specific to the Present Model**

An issue which assumes a central position in this model is that of the personality or style of the clinician. Anything which affects the clinician's openness to his own inner experience or his perceptiveness of subtle interpersonal cues must affect his ability to make judgments about the experience of others.

A number of questions might be asked concerning the growth of the internal model itself; the rate it is formed, external and internal factors influencing this rate of growth, differences in the level of consciousness at which access to the model is possible, etc. Interesting methodological problems are raised by the necessity of translating "formation of an internal model" into operational terms. One approach would be to use successive Q-sorts. Another approach might be to obtain some measure of the clinician's "stereotype" of a person of the subject's class (average patient, average college student, etc.), then define the formation of a model specific to this subject in terms of the clinician's modification of this stereotype.

Explorations based on this model of clinical judgment could have consequences for training in clinical psychology. Training which enables one to make accurate judgments about the inner state of another person is different from the training which enables one, for example, to classify MMPI profiles as neurotic or psychotic, but research which has attempted to influence clinical judgment through training has been concerned with judgments of the latter type. Perhaps greater awareness of the differences among types of clinical judgments and increased study of types previously neglected will permit training programs to be designed in accordance with the desired final product.

---

Condensed with permission from the *Bulletin of the Menninger Clinic*, 36(3):279-301, 1972. Copyright 1972 by the Menninger Foundation. Reprints of original article from Michael K. Harty, Ph.D., Staff Psychologist, The Menninger Foundation, Box 829, Topeka, Kans. 66601

# Personality and IQ Measures as Predictors of School Achievement

K. Barton, T. E. Dielman, and R. B. Cattell

Many studies have related measures of IQ to measures of achievement in school. On the average, such studies indicated a correlation of about .70 between IQ and achievement. As this figure shows that only about 50 percent of the variance in achievement scores can be accounted for by IQ, investigators looked elsewhere for other determining variables, and attempts were made to link the prediction of achievement with personality variables.

This study was designed to assess more fully the relative importance of ability and personality variables in the prediction of school achievement. The hypotheses to be tested were (*a*) by using a combination of intelligence and personality variables to predict school achievement, better predictions would be obtained than when intelligence variables were used alone; and (*b*) achievement in different subject areas will be predicted best by a set of personality and intelligence variables unique to each area.

The subjects were 169 sixth-grade and 142 seventh-grade students. Socioeconomic status ranged from upper-middle class to lower class in about equal proportions. All subjects completed the Culture Fair Intelligence Test and the High School Personality Questionnaire in January 1970. In addition, four standardized achievement tests in mathematics, science, social studies, and reading were given in March 1970, thus allowing a 2-month interval between the predictors and the criteria. The High School Personality Questionnaire measures a set of 14 factorally independent dimensions of personality shown in Table 1.

All tests were computer-scored and correlations obtained among the personality, ability, and achievement variables. Multiple-regression analyses were performed to obtain measures of effectiveness of each variable and different combinations of variables in the prediction of school achievement. The regression analyses were designed to test several hypotheses relating to the predictive value of different intelligence and personality variables:

Model 1 (Culture Fair Intelligence Test + High School Personality Questionnaire versus Culture Fair Intelligence Test) was designed to indicate whether an increase in prediction of achievement was made when the whole High School Personality Questionnaire was added to the Culture Fair Intelligence Test results.

Model 2 (Culture Fair Intelligence Test + High School Personality Questionnaire versus Culture Fair Intelligence Test + B) indicated whether the addition of just the nonintelligence factors in the High School Personality Questionnaire increased achievement prediction over that obtained when both the Culture Fair Intelligence Test and Factor B were used.

Model 3 ([Culture Fair Intelligence Test + (High School Personality Questionnaire − B)] versus Culture Fair Intelligence Test) allowed for the assessment of the nonintelligence factors in the High School Personality Questionnaire over and above the effectiveness of the Culture Fair Intelligence Test alone.

Model 4 (Culture Fair Intelligence Test + B versus Culture Fair Intelligence Test) showed the effectiveness of adding Factor B to the only other intelligence measure, the Culture Fair Intelligence Test.

Models 5, 6, 7, and 8 indicated, respectively, the effectiveness in prediction of achievement of the High School Personality Questionnaire alone, the High School Personality Questionnaire nonintelligence factors alone, Factor B alone, and the Culture Fair Intelligence Test alone.

The analyses show that factors sig-

Table 1. Personality Factors Measured by the High School Personality Questionnaire

| Low-score description | Alphabetic designation of factor | High-score description |
|---|---|---|
| Reserved | A | Warmhearted |
| Low intelligence | B | High intelligence |
| Emotionally unstable | C | Emotionally stable |
| Undemonstrative | D | Excitability |
| Submissive | E | Dominance |
| Desurgency | F | Surgency |
| Weak superego | G | Strong superego |
| Shy | H | Socially bold |
| Toughminded | I | Tenderminded |
| Zestful | J | Reflective |
| Self-assured | O | Apprehensive |
| Group Dependency | $Q_2$ | Self-sufficiency |
| Uncontrolled | $Q_3$ | Controlled |
| Relaxed | $Q_4$ | Tense |

nificantly related to all four measures of achievement in both grades are the two measures of IQ (Factor B of the High Personality Questionnaire and the Culture Fair Intelligence Test) and Factor G of the High School Personality Questionnaire, a measure of conscientiousness.

In the sixth grade, Factor A (warmhearted participation) is significantly related to all four measures of achievement. In addition, Factor H (adventurousness) is significantly related to achievement in mathematics and Factor $Q_4$ (anxiety) is significantly related to social studies.

In the seventh grade, Factor C (emotional stability) is significantly correlated with mathematics, reading, social studies, and science. Factor E (dominance) is significantly related to mathematics. As in the sixth grade, Factor H is significantly related to mathematics. Factor I (toughmindedness) is significantly related to mathematics and science and Factor J (desire for group action) is significantly related to mathematics and science and to social studies. Factor O (self-assuredness) is significantly related to all achievement measures as is Factor $Q_3$ (exacting willpower).

In both the sixth and seventh grades, Model 1 indicates that the addition of the High School Personality Questionnaire to the Culture Fair Intelligence Test results in a significant increase in the amount of variance accounted for in achievement in all four areas. Model 2 indicates that in the sixth grade, adding the nonintelligence factors of the High School Personality Questionnaire to the Culture Fair Intelligence Test and Factor B results in increased power of prediction for reading only. However, in the seventh grade this is true for both reading and mathematics. Model 3 indicates that even if only the nonintelligence factors are added to the Culture Fair Intelligence Test a significant increase in variance is obtained in the sixth grade, but this is only true for mathematics and reading in the seventh grade. The significance of just using the nonintelligence factors of the High School Personality Questionnaire alone is given by Model 6; in both grades and in all achievement areas this is significant. The effectiveness of the whole High School Personality Questionnaire is revealed in Model 5; it is highly significant in both grades and for all subjects. Models 7 and 8 show that in all grades and subjects both measures of intelligence predict achievement. The Culture Fair Intelligence Test is superior to Factor B in the seventh grade for mathematics, whereas Factor B is preferable in the sixth grade for both social studies and science. Model 4 demonstrates that

adding Factor B to the Culture Fair Intelligence Test results in a significant increase in prediction for all subjects and grades. This result is not surprising as the Culture Fair Intelligence Test was designed to measure a different form of intelligence (fluid or innate) than was Factor B (a more crystalline measure influenced by environmental factors).

Finally, several conclusions can be made regarding which tests best predict achievement in social studies, science, math, or reading.

*Social Studies*—Measures of both IQ and personality are superior to either type of measure taken singly. However, a comparison between Models 1 and 5 indicates that the High School Personality Questionnaire taken alone (as it contains a measure of intelligence) is as effective as when it is taken in addition to the Culture Fair Intelligence Test.

*Science*—The results are the same as for social studies, that is, the High School Personality Questionnaire provides an adequate prediction instrument without the Culture Fair Intelligence Test.

*Mathematics*—Here the Culture Fair Intelligence Test measure considerably improves in its predictive value in both grades, but personality variables (Model 6) are important especially in the seventh grade. Model 1 compared to Model 5 indicates that in this area the best prediction of achievement can be obtained by using both the High School Personality Questionnaire and Culture Fair Intelligence Test measures.

*Reading*—The results are the same as for mathematics, that is, IQ and personality measures are both important and the Culture Fair Intelligence Test + High School Personality Questionnaire provides for the best prediction of achievement.

In brief, three main conclusions can be made: (1) There seems to be a general concept of achievement consistently related to a set of personality and intelligence measures over all four achievement areas. This set of variables includes IQ and Factors B and G. (2) There are personality factors which follow a developmental sequence in their relationship to achievement. (3) Some personality factors are specifically related to certain academic subjects.

---

Condensed with permission from the *Journal of Educational Psychology*, 63(4):398–404, 1972. Copyright 1972 by the American Psychological Association, Inc. Reprints of original article from Keith Barton, Ph. D., Department of Psychology, University of Illinois, 531 Psychololgy Building, Champaign, Ill. 61820

# A Nonverbal Semantic Differential

**P. M. Bentler and Allan L. Lavoie**

The semantic differential has been used in cross-cultural research and in developmental studies. Both these areas might profit by the development of a nonverbal measure of semantic space. Such a measure could eliminate the problems associated with finding translation equivalents for adjective scales. A nonverbal measure would also enable group testing of much younger children, would allow testing of nonliterate groups, and might be useful with aphasics.

Little has been done toward developing a reliable nonverbal measure of Evaluation, Potency, and Activity (E, P, and A). We attempted to find graphic scales that would measure EPA as well as the recently discovered dimensions Density, Orderliness, Reality, and Familiarity (D, O, R, and F).

### Experiment I

The subjects were 320 undergraduates. We had 160 subjects draw abstract figures in response to stimulus words. The stimulus words were items which defined the seven verbal semantic differential factors. Each subject drew 10 figures, and for each word at least six figures were drawn by different subjects. Figures drawn in response to "good" were paired with figures drawn to "bad" and so on, so that each bipolar pair of adjectives was matched with bipolar figures. From these bipolar pairs of figures, we selected the clearest and least ambiguous. The result was 26 relatively clear pairs.

The second stage proved necessary when we found that two factors had no clear pairs (Reality and Familiarity) and two others had only one each (Evaluation and Activity). Orderliness was overrepresented with 17 pairs, Potency had four, and Density three. We generated new items to identify all dimensions except Orderliness, which was adequately defined. The result was 70 items. Each dimension was represented by at least six item pairs. The figures were placed in a forced-choice format, 10 pairs to a page, with order randomized.

Forty concepts were selected from previous semantic differential studies. Two males and two females rated a randomly assigned concept. This produced 160 concept ratings made by 160 different subjects. They were directed to choose that member of each pair which better described their concept.

The 70 figure pairs were intercorrelated across subjects and hence across concepts. The factor matrix was rotated using Clustran, a transformation procedure that rotates an initial loading matrix to a hypothesized target. Seven factors were hypothesized corresponsding to EPADORF. Reality and Familiarity failed to come out, while E and A were identified only by positively arranged items. Orderliness was identified by 46 item pairs with loadings greater than .35.

The original data were not further analyzed for several reasons. First, almost every item had some loading on Orderliness. Second, it appeared that item pairs were being strongly affected by response styles. And third, the failure of any of the R and F items to emerge independently of Orderliness was a further indication that the subjects were using the figures concretely. It looked as if the subjects were using the stimulus characteristics of the items to match the stimulus charac-

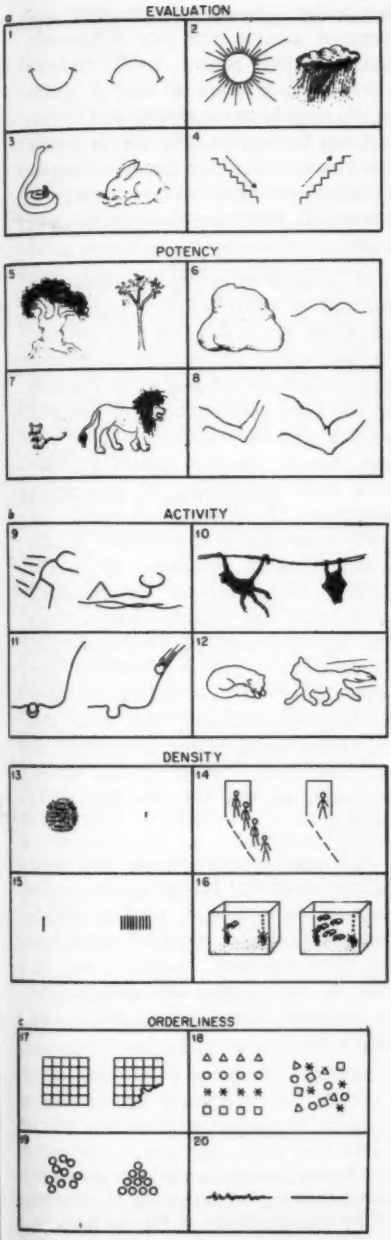


FIG. 1. Representative nonverbal items.

teristics of the concepts, rather than responding in terms of meaning.

**Experiment II**

The subjects were 160 undergraduates. This experiment involved rating the same concept using the verbal scales that the subject rated using the graphic scales.

We decided that the best way to suppress response styles and stimulus-matching responses was to use items that represent the relevant meaning dimension more specifically and less abstractly. For example, for E we drew a heart vs. a dagger to represent "kind-cruel," a boulder vs. a pebble for P, a river vs. a pond for A. The result was 118 items, including the 70 original. The concepts were the same ones used in Experiment I.

Two males and two females again rated a randomly assigned concept. To allow clearer identification of the graphic factors, the subjects' adjective ratings of the concepts were factor analyzed and rotated to the verbal EPADORF. Their figure ratings were included in the data matrix but the nonverbal scales were permitted to align themselves relative to the seven verbal factors. From this analysis, the nonverbal scales loading most clearly on the verbal factors were used as pivots for a nonverbal analysis alone. It was possible to select items which could be used to hypothesize five nonverbal factors. The best nonverbal items for each factor were thus identified. There were 75 acceptable items.

Finally, total scores were obtained for each verbal and nonverbal factor by adding the respective items having high loadings in unit-weight fashion such as when scoring a test.

Five factors emerged. These were easily identifiable as EPADO. As in Experiment I, the R and F items

loaded primarily on Orderliness, and O had the largest number of items loading greater than .35.

To assess the degree to which the nonverbal EPADO were measuring the same dimensions in the verbal ratings, total verbal scores on each factor for each subject were correlated with the corresponding nonverbal total scores. The average cross-method correlation was .63.

The results indicate that at least five of the verbal factors, EPADO, can be measured reliably nonverbally and that the resulting factors are relatively independent. O correlates fairly high with E, but correlation is not too high to disqualify O and E as separate factors, especially when compared to their internal consistencies. But R and F failed to emerge. It was decided to try once more.

**Experiment III**

The subjects were 160 undergraduates. Forty-five new figure pairs were developed for R and F, and to increase the size of A and D. This brought the total to 120. The same concepts as in Experiments I and II were used, and the procedure was the same as in Experiments I and II.

Again, five clear factors emerged, the same as in Experiment II—EPADO.

Whether there are two more nonverbal factors corresponding to verbal R and F is unanswerable by the studies reported here. There are several reasonable explanations for their failure to emerge. It may be that our concept sample was not sufficiently variable with respect to R and F; that our nonverbal R and F scales really were better measures of O; that subjects have great difficulty in visualizing real vs. unreal, familiar vs. unfamiliar; or that visual representations of R and F are used idiosyncratically.

A second question concerns the method by which subjects make their decisions. Some subjects may primarily use visual imagery in something akin to sensory processing. If subjects in the nonverbal semantic differential studies used the visual mode of thinking, that is, stimulus matching, there is a high probability of denotative matching. If, on the other hand, our results are due to conceptual or abstract thinking processes, where the subject first abstracts the meaning of a scale, the chances seem better that affective meaning is being tapped. If the latter is true, it is unlikely that very young children would be able to perform similarly, and thus the nonverbal scales would be limited to children about 7 years old and older. Research on the response process is needed.

A further question is the relevance of the nonverbal factors to the problems of cross-cultural studies. While research is being planned, none has yet been executed. However, the results of the studies are encouraging in terms of factor independence and reliability.

---

Condensed with permission from the *Journal of Verbal Learning and Verbal Behavior*, 2(4):491–496, 1972. Copyright 1972 by Academic Press, Inc. Reprints of original article from Peter M. Bentler, Ph.D., Assistant Professor, Department of Psychology, University of California, Los Angeles, Calif. 90024

# The Psychodynamics of Examinations

**M. L. Moeller**

In the psychodynamics of examinations, there are three principal parts: the psychodynamics which lead us to the psychosocial phenomenon—the ritual of the examination situation within society, especially with universities; the psychodynamics between the testee and the testor as well as the interactional dynamics between the two under examination conditions; the psychodynamics that take place when the examination situation is changed, reformed, or is to be revolutionized.

### Examinations as Social Rituals

In modern society, examinations are instruments of cultural selection; they decide the course of life for the individual in our society. There is little that we can say about the usefulness of examinations without considering the genetic psychodynamics. We must ask ourselves how it was possible for the phenomenon of examinations to maintain its status as an interactional process and as a ritual within the socialization processes. Only when this is determined can we investigate deeply the examination system as a psychosocial symptom within a special sociopsychological sector, particularly within the university dynamics of a society.

As long as the unconscious meanings of examinations have not been identified, examination reform will never result in emancipation. Instead, it will always be guided and steered by the unconscious impact of an examination. Language itself will give us a few hints toward this unconscious impact of examinations. Graduation is regarded as a test of maturity and accomplishment; therefore, it points not only toward intellectual but also toward emotional and instinctive aspects of man.

In tracing back the history of examinations, earlier peoples attempted to find the solution to a psychosocial problem by puberty rituals which involved a decisive, intense change in the mutual relationship between generations. This change is seen unilaterally by the young people who are assimilated with equal rights into the society of adults, but for the grown-ups there also ensues a role change—forsaking the parental position vis-à-vis that of the young person. In other words, the relationship between testor and testee corresponds to a parent-child relationship in a critical phase of reversal. Such conflicting interactional processes, in which both sides sides are in jeopardy and a correspondingly intensified fear is generated, beguile groups by way of ritualization. The examination is based on one such ritual which may be regarded as the social execution of the development of the instincts in puberty. The parent-child relationship must be converted into a symmetric role-relationship, and the ritual takes on different forms depending on the social structure. At the heart of it, the ritual reflects the corresponding relationships in society—the shaping of societal life, the formation of symbols, and the existing living conditions.

The concept of a ritual always includes mutually supportive roles. Conflicting and often debilitating ambivalence is surmounted through the compromises of individual demands, that is, through inevitable frustrations. And so we see that the basis of an examination is a ritual in which

society attempts to deal with a new situation and with its related ambivalences, that is, with a change in a reciprocal relationship and with a conflict that arises from it. The ritual is the attempt to reduce conflicts and anxieties. It is, in itself, neither good nor bad; it is simply the attempt to surmount conflict.

The examination is arranged to induce fear. It not only generates individual fear symptoms, but it causes fear itself. The ritual has become an asymmetric attempt to overcome conflict in which one party remains unscathed at the expense of the other. It permits the dominance on the part of the parents, thus corresponding to the authoritarian structure of our educational reforms. Our examination has degenerated into an interactional symptom whose principal characteristic is the asymmetry that yields an uncontrolled power within the ritual to the parent generation.

The effect of this asymmetric examination structure is that the fear and anxiety generated in the ritual is predominantly on the testee. The intensity of this fear changes generally with the various forms of examinations and the different faculties. All investigations show that fear has had a negative effect on the capacity of proof in examinations; consequently, the examination prevents the very process it is supposed to measure. In today's form, the fear it mobilizes makes the examination the most unfit instrument conceivable for measuring accomplishment.

Now we must add yet another paradox. It is true that an examination not only measures, it also evaluates. Psychodynamically speaking, the measuring function is supported by the function of a status assignment. Testees are either recognized or acknowledged as to their knowledge and social role or they are not so acknowledged; the examination expresses a social existence or it does not. That fear inhibits not only the efficiency of the testee and the judgment of the testor, it affects everyone who has anything to do with the examination. It seems as if the examination situation would admit the most intelligent interpretations; however, emotions determine the whole process of the examination. We are concerned here with the central irrational effect of the examination—with an unconscious emotional taboo corresponding to the suppression of the very fear that the examination awakens in all of us. But since the examination brings about this fear by its construction, which creates advantages for the parents, it must be assumed that power relationships are stabilized in this manner. Examinations as a sequentially graduated selective process reduce the portion of low achievers from one educational level to the next and stabilize the middle-class culture in the academic field. From a psychoanalytic viewpoint the induction of fear through examinations invites suppression and thus submerges the irrational characteristics of the situation into the unconscious. Withdrawing irrationalities from criticism, it permits reforms to be converted to repetition, that is, to the perpetuation of the situation in a different form.

### The Pschodynamics of the Examination Event

The psychodynamics of the actual happening is based on the triad which includes examinee, examiner, and examination arrangement. But the three factors do not exist by themselves. It is important to note the interaction between the examiner and the examinee and how much the interactional dynamics are altered through the vary-

ing arrangements of the examination.

The examination ritual marks only one developmental step that has been preceded by many others and that will be followed by others during the socialization of the testee and of the testor. The examination, then, must fit into this social-specific series of rituals. The less it is attuned to this series of rituals, the more inadequate it becomes and the stronger is the fear that it mobilizes in the individual. And, if it is completely excluded from the developmental framework, psychodynamically speaking, the apparent benefit of freedom from fear and the elimination of repression would not occur as one would expect from a purely rational perspective. Instead, social recognition in education or job-training—this mutual approval which is an essential element in identity formation—would be entirely lost.

Psychodynamically, then, the lack of interaction would compare to isolation of the group, which means the individual could not enter that community but would have to shift for himself. It is likely that the most difficult examinations would not mobilize such a profound fear as that fear which is created when a person is not being examined at all. No man can be deprived of some form of social recognition without experiencing fear. The problem seems to be to transform the useless, distorted ways of examining into a mutual acknowledgment of essential developmental steps. It must be remembered that examinations and the tensions they produce may be part of normal development, and we are not seeking freedom from any sort of fear, but freedom from irrational fear. The functions of examinations and of examination fears as optimum frustrations in the service of the socialization process were among the aims of the original ritual for both participants.

*Unconscious dangers experienced in examinations*—Clinical investgations of students frightened by examinations have yielded four dangerous situations which the testee can experience unconsciously in modern-day examinations. The fear of examinations is the reaction to these unconscious dangers originating in childhood history. Since frequently a number of dangerous situations are experienced simultaneously, the fear of examinations is a mixture of anxieties.

*Separation*—One of the earliest dangers for the child who is still helpless is separation from his parents. The examination as the conclusion and demarcation of a certain period of education is regarded by the unconscious as a danger rendered manifest by the experience of the separation. For instance, a woman student suffered from fear and panic because she was threatened with losing the motherly protection of the university from which this examination was about to separate her. Personalities with depressive tendencies develop most severely this fear of being forsaken in examinations.

*Offense*—Another developmental disturbance leads to an unconscious conception of one's own omnipotent greatness; therefore, any situation in which such a mentality is validly confronted with reality must be experienced as a danger. In an examination situation, the danger of not feeling this narcissistic omnipotence confirmed in intellectual prowess evokes great fear.

*Punishment*—In an examination situation, the examinees experience some sort of a day of final judgment in which they must atone and be punished for unconscious past sins,

that is, for unconscious libidinous or aggressive wishes. The examiner, as representative of the parent figure, takes over the function of the conscience in his role as the administrator of punishment. The work to be accomplished becomes a passion understood as the expression of unconscious motivations; the confrontation with the examiner brings to life the forbidden wishes. Punishment may also be desired masochistically. That is, it can become a temptation to damage oneself, for instance, with testees who cannot shake loose the unconscious desire to fail in the examination or to commit a subtle sort of suicide on the job.

*Temptation*—This situation becomes obvious when the student imagines killing the examiner as his father—an aggressive temptation—or when a woman student dreams of sexual intercourse with the examiner—a libidinous temptation. The significance that students assign to their antipathy or sympathy toward the examiner might be regarded as a reflection of these deeper instincts. But an even greater conflict is caused by the temptation which operates in every human being, namely, the desire for his opposite sex role. Men are afraid of their own passive desires, namely, the homosexual temptation in the examination, and women of the fact that they are not evaluated as men are. Rivalries between siblings that achieve a pathogenic dimension in childhood lead to a special aggravation of the normal competitive situation in an examination, and with that, to a special form of aggressive temptation.

These four unconscious dangers experienced by examinees can be summarized as "dangers of identity." They depend not only on the personality of the testee, but also on other factors. For the opponent, the testor, also has a history that is, according to expectations, neither more nor less pathogenic than that of the testee. The decisive difference is that the testor has power with which he actually determines the fate of his subjects. This power, given the same unconscious conflicting problematics, leads to a reversal of the behaviors. When fear is supplemented with power, it turns into an authoritarian behavior. Just as in the case with the testees, the behavior of the testor can easily be modified through the four basic situations. Testors who see in the examination a threat of separation will attempt to make the separation more difficult, i.e., passing that examination. They behave more like mothers who are reluctant to let go of their children. This inhibitive behavior, not a conscious act by the testor, increases fear in certain testees.

Even better known is the behavior of narcissistically disturbed examiners who regard the examination as a possible situation for personal devaluation. The examination is well-suited to exploiting this tendency, since usually the material for the examination concerns the personal field of endeavor of the testor; the ignorance of the testee, therefore, can be experienced as a personal insult. This personality trait of the examiner again increases the fear of the examination. Sadistic methods of testing may serve as a defense against a punishment situation, but sadistic behavior can also be generated from the mere temptation to permit aggressive influences to run amok. Such sadistic authoritarian behavior occurs unconsciously and plays a great role in the fantasies of testors. Given the same unconscious experience of the examination, the examiner may succeed in saddling his weaker partner, the testee, with his

own fear through the power that the situation has given him. By so doing, he augments the testee's fear of the examination. The unconsciously authoritarian posture can also generate strong guilt feeling in the testor and thus lead to a diffusely soft, pseudoliberal attitude of the testor that would accept anything. But the mental conflicts of the examiner not only lead to his unburdening fears on the testee, they also distort his own ability to render judgment.

### Irrational Qualities of Examinations

Now let us consider the third major factor of the examination: the examination regarded in the light of primitive, irrationally institutionalized elements. Even today almost every examination is a unique situation. It is earmarked by an excessively high significance of one single examination, by transparency of the examination, by unchallenged power of the testors, and by a uniformity in format which, because each examinee has his own individual working style, taxes each one differently and is therefore unjust. This feature of being unique is made even more acute through the nonrepeatability of the examination. A similar phenomenon is the coupling of serial examinations whereby failing a specified number of the examinations would necessitate repetition of the entire series of examinations. The centralization of such irrational characteristics must frighten any examinee so that to be unafraid of examinations today is more likely to be regarded as pathological than to be afraid of examinations.

The most irrational characteristic of all in regard to examinations is that these irrational characteristics remain unnoticed by almost everyone, and that they are accepted as matters of course. It is surprising that in a rational society such irrational processes of judgment can still prevail in grades, esteem, and credibility, right down to affecting decisions such as the giving of grants and occupational employment. The evaluation of 1 single hour is purported to be representative of the accomplishments in a course of studies that encompasses several years. From a psychodynamic viewpoint, today's examinations have the value of scrap. The assignment of a grade is an evaluation of the particular relationship between the testor and the testee, a highly specific symptom of their interhuman relationship which hardly lends itself to generalization.

### The Psychodynamics of Examination Reform

First I should like to summarize what might be underestimated in reforms: the influence of the emotions on the intellectual accomplishment of the testee and on the judgment of the testor; the dependency on such a ritual, in spite of fear, because it represents overcoming a conflicting change in the relationship of generations by the result of a new mutual recognition; the irrational characteristics of the examination that induce fear; the significance of the unconscious conflict in the testee as well as in the testor; the fact that every phenomenon in the examination cannot really be understood without considering its interdependent effect in the total overall psychosocial constellation.

If we now regard the whole reform effort, we discover how very much the new developments are being guided by the manifestly unconscious impact of the examination. Based on the emotional taboo, the reform inadvertently runs the danger of propagating forced repetition. I

should like to cite a few examples:

The fractionation of examinations is desirable, because it eliminates the massing of tests into a large final examination after years of study. The fractionation of the examination into small tests accompanying the studies has converted the entire course of study into a permanent examination situation in which one is never free from stress.

The so-called objectivization is the attempt to render the accomplishment of an examination objectively measurable. Here the fact is overlooked that more objective values are rendered, but that these still represent uncertain accomplishments. Furthermore, all of the rest of the irrationality of today's examinations remain unchanged, to say nothing of the questionable value of the written system for some subjects.

Since graduation demands vary according to college and even according to state, and since such inequality is unfair, any sort of standardization in the sense of equality seems desirable. The central error, however, lies in the negation of the individuality of learning and accomplishment. Objectivization and standardization, then, are inequality for the individual, because everyone's method of achievement is different. It would be dehumanizing to permit examination arrangements that do not consider these individual peculiarities.

What suggestions for reform can be derived from the psychodynamic perspective? I believe the most worthwhile contribution would be the recognition of the emotionality and the individuality of learning and accomplishment. It would include: A pluralism in the methods of examination encompassing forms which, in today's sense, would not be called examinations; freedom of choice for the testee between these multiple methods of examination; elimination of all irrational aspects of today's examinations; emancipation of the content of examinations and the methods of examinations by equal participation of testors and testees who have equal rights to suggest changes.

Today, and even more so in the future, we find ourselves in a scientific explosion in the sense of increasingly rapid changes in the structure of knowledge. Because of this rapid change in knowledge, permanent examinations are threatening. Learning, or more precisely learning to learn, is an important factor in the human identity. A ritual that serves this identity formation is therefore well-suited to transmitting knowledge and experience between the learner and the teacher—if it does not petrify in sanctified untouchability—but if instead it undergoes emancipatory changes to reflect a flexible mutual relationship.

---

Translated and condensed with permission from the *Zeitschrift für Psychotherapie und medizinische Psychologie*, 22(1):1–13, 1972. Copyright 1972 by Georg Thieme Verlag Stuttgart. Reprints of original article from Priv.-Doz. Dr. Michael Lukas Moeller, Psychosomatische Klinik der Universität Giessen, 63 Giessen, Ludwigstrasse 76, Germany.

# Transitions in Infant Sensorimotor Development and the Prediction of Childhood IQ

**Robert B. McCall, Pamela Savoy Hogarty, and Nancy Hurlburt**

The purpose of this article is twofold. First, the literature concerning the prediction of childhood and adult IQ scores from infancy is reviewed, focusing on the possibility of psychometric factors clouding long-term prediction. The other goal is to consider an alternative conception of intelligence, to describe its theoretical nature, and to illustrate empirically its possible character.

### Prediction from Infancy

*Normal children*—It is an overstatement to say that there is "no prediction" from the infancy period to childhood IQ. There are correlations that attain statistical significance between certain ages for some tests, but the size of the relationships is not particularly impressive and certainly not adequate for clinical application.

Data pooled from several studies show two obvious trends: When predicting to any age in childhood, the later the test is given during infancy period the higher the relationship. Correlations to IQ at 3–4 years are higher than with IQ scores assessed thereafter.

*Nonnormal samples*— One purpose of developing infant tests was to detect children who might have severe pathologies. Several studies have correlated total score on infant tests with later IQ for samples involving subjects having extremely low infant test scores or subjects having known or suspected abnormalities. A very low score on an infant test, even though a pediatrician does not classify the child as abnormal, may have diagnostic value.

While correlations for suspect and clinical populations are high, they must be applied with caution. Often, correlations in the range of .70-.80 for special clinical populations are cited as justification for use of the assessment instrument to screen more randomly selected populations. This is not always practical because of the difference in base rate of abnormality in the clinical versus the random sample.

Caution should also be invoked on theoretical grounds. The correlations reported for clinical samples may merely reflect the presence of some subjects whose deficiencies are so marked that there is practically no behavior to assess at either age. Such low scores by "default" can produce sizable correlations in an otherwise variable sample.

*Reliability of infant scales*—Few people question the reliability of childhood IQ scores, since test-retest correlations rarely fall below .85. In contrast, many people suspect that the infant tests are unreliable, but this does not seem to be the case. While infant scale reliabilities are not particularly high during the first 3 months, the internal consistency and the test-retest comparability of infant scales after 3 months of age are substantial (high .80s and low .90s). Since the infant tests are approximately as reliable as the childhood IQ assessments, reliability does not appear to be a major factor in the failure to obtain prediction from infancy to later intelligence.

*Sex differences*—While sex differences are not usually found for the general scores from the Bayley or the

Gesell tests, this fact does not obviate differences in cross-age correlations for the sexes. If the sexes differ in the nature of the relationship between infant performance and later IQ, combining the sexes into a single sample might prevent these relationships from appearing.

Studies of sex differences in predictability yield conflicting results that are difficult to interpret. The studies differ in sample characteristics, national origin, infant and childhood tests, etc. Nevertheless, the data are consistent with the suspicion that Gesell predictions to childhood IQ may be higher than for other infant tests and that girls may be more predictable from the first year of life than boys, but long-term predictability from later in the infancy period appears to be comparable for the two sexes.

*Socioeconomic status*—There appears to be some justification for considering other variables in concert with the infant test when attempting to predict later IQ. Specifically, for normal infants parental education or socioeconomic status is probably a better single predictor than the infant test during the first 12–18 months of life. Thereafter, the level of prediction may be elevated slightly by adding the infant test score into a multiple regression with socioeconomic status, and the multivariate advantage may occur at an earlier age for girls than for boys.

*Specific items and abilities*—Another approach to the problem of prediction from infancy is based on the proposition that specific subgroups of items within the test might have greater predictive saliency than the total score. If many of the items on an infant test are irrelevant to the future course of cognitive growth, they would contribute "noise" to the total score which would mask the effect of items that actually are predictors of later mental performance.

Studies appear to show that specific items and abilities might be more predictive of later IQ than the total score of an infant test and that there may be sex differences with respect to which items and skills are most predictive. The issue is whether there is any consistency in the general nature of items that predict later mental ability versus items that do not.

The most replicable item subset which predicts from infancy to later IQ involves early vocalization behavior, and the predictions are found more frequently for girls than for boys. It is unlikely that this sex difference is merely a simple reflection of precocious verbal behavior on the part of females relative to males. If this were true, then the age-of-first-passing scores would merely be a few months younger for girls than for boys and such a difference in the means of these distributions should not influence the correlation. Thus, there may be some qualitative sense in which early vocalization behavior has special meaning for girls.

### Basic Concept of Development

The search for correlational stability across different ages implies faith in a developmentally constant, general conception of intelligence ("g") that presumably governs an enormous variety of mental activities. Under that assumption, the nature of the behavioral manifestations of "g" would change from age to age, but g itself is presumed constant, and thus mental precocity at one age should predict mental precocity at another. Confronted with the evidence, this g model of mental development must be questioned.

It is quite likely that the several skills that characterize a given stage

in infancy will vary considerably in the extent to which they share common determinants. Thus, at any specific age there might be some skills that are highly related and others that are independent. This same proposition may apply when skills are viewed developmentally. Consequently, one way to approach the study of mental development in the first several years of life is to consider the matrix of possible Skills x Skills x Ages (ideally, developmental levels), and to ask not only which skills are related to or independent from other skills within an age (Skills x Skills) but which skills are related to which other skills across ages. The cross-age correlations mark one type of developmental transition.

The belief in a general mental ability assumed that there was some constant determinant that governed the performance of a vast number of skills both within and across ages, and thus one should expect high within-age correlations among diverse behaviors and a single, unified, and continuous trend of relationships across age. The fact that many different abilities of childhood did show positive correlations presumably testified to a single underlying dimension which then became reified under the rubric of intelligence (or g). Soon, intelligence was used to explain the very behavior that mothered its theoretical existence. The evidence for a single major mental ability either within an age or across age is not very convincing. However, neither do these data argue for the conceptually opposite conclusion. It seems unlikely that there are no common processes underlying some mental activities, no hierarchical organization of skills, no generality from one task to another, and no predictability from one age to another for some processes. The Skills x Skills x Ages matrix of correlations is neither homogeneously 1.00 nor .00.

Consequently, one model of mental performance with heuristic value might fall somewhere between these two extremes. First, one would suppose that within any given age (or developmental level) one could discern subsets of skills that would be highly related within a set, but the subsets would be independent from one another. Some of these skill sets might possess relationships with similar or qualitatively different skill sets at a future age; others might lead to a "dead end" in terms of their developmental significance. The matrix of developmental transitions and interactions might be quite complex.

The remainder of this essay is devoted to describing empirically such possible transitions during infancy and childhood with the purpose of illustrating this conceptual approach and generating hypotheses about the fabric of early development.

#### Developmental Transitions

The responses of infants to individual items on Gesell tests at 6, 12, 18, and 24 months of age were subjected to principal component analyses separately at each age. The resulting components presumably represented independent skill areas within each age. Next, the component scores were correlated across age (and with childhood IQ scores) and the network of developmental transitions examined.

*Subjects and analyses*—The subjects were all children who had at least one Gesell test at either 6, 12, 18, or 24 months, and who had at least one of the following assessments in childhood: a Binet at 3½, 6, or 10 years; the WISC at 7 or 11 years; or the Wechsler–Bellevue at 13 years. Separate component analyses were performed for the two sexes. While

the component structure within each age did not appear different for the sexes, the correlations between component scores across age did contain sex differences.

*Results*—The component scores at each age were correlated. The correlation matrix depicted several general characteristics: First, each sex has a single main developmental trend consisting of the items loading on the first principal component at each age. The main trend was the only developmental pattern in which adjacent-age and nonadjacent-age correlations were both significant. While this major trend consisting of the first principal components has the appearance of a g factor, none of the components accounted for more than 19 percent of the total test variance.

Second, at 6 months the items in the main trend do not predict childhood IQ, though some smaller components do correlate with childhood IQ. This observation has several implications. Initially it suggests that precocity in certain skills characteristic of early infancy may not relate to later childhood IQ, even though such skills may be vital precursors of those childhood traits. Further, since the main trend was composed of the first principal component at each age and since that component is presumably the best single concentrated reflection of the test as a whole, a great deal more information and predictive power may lie within subsets of items than is possessed by the total score.

A third point is that the two sexes were almost totally different in the pattern of relationships across the infancy period, except for the main trend. This observation highlights the importance of the main trend in infant development but also indicates that the sexes may follow some distinctly different paths in the course of early mental development.

Fourth, the analysis suggests that the important developmental aspects during infancy are somewhat more concentrated in the main trend for girls and more fractionated and dispersed among other behavioral characteristics for boys.

*Interpretation*—The interpretation of a component analysis is necessarily a subjective enterprise. Consequently, the interpretation given below is offered for heuristic reasons and should be viewed as one of many possible implications of these descriptive data.

The items loading on the main developmental trend at 6 months appear to represent visually guided manipulation of objects and gross and fine motor behavior. Many of these behaviors involve the manipulation of objects that yield perceptual contingencies. For example, the items "lifts inverted cup," "bangs spoon on table," "splashes in tub," and "pats table" all describe manipulation that produces a rather clear contingent perceptual consequence. With due interpretative caution noted, the "perceptual consequences" hypothesis is a particularly exciting one because of its relationship with other behaviors that are similar to it. For example, it has been shown that infants as well as animals will perform simple responses in order to produce a contingent perceptual event. Also it has been suggested that reciprocal behavior and social-perceptual contingencies may form the basis of early attachment in the human infant. Finally, one is tempted in this context to think of the concept of internal control found important in later intellectual, educational, and personality domains.

At 12 months, the items composing the main developmental trend still involve gross and fine motor skills,

but the items reflecting visually guided manipulation and production of perceptual consequences seen at 6 months have been replaced by items requiring the imitation of fine motor behavior and the learning of rudimentary social skills. Also loading on this component are several diverse behaviors that seem to reflect the learning of social behavior and simple verbal skills. The presence of these social-verbal behaviors on the same component with the imitation of fine motor behavior suggests that these are related behaviors, perhaps tied together in a social context. It seems reasonable that the child who develops the tendency to imitate probably imitates fine motor behavior first and later vocal-verbal skills in social contexts.

By 18 months, the main developmental trend has become pervasively verbal in character, with three general types of items loaded on the first principal component: verbal production, verbal comprehension, and an extension of the imitation of verbal and motor behavior. The propensity to imitate fine motor behavior and verbal skills in a social context at 12 months is related to the imitation of more complex verbal behavior and skill in verbal labeling and comprehension at 18 months.

At 24 months, the main trend has an even stronger verbal character. While there is still a hint of imitation, the predominant theme features the verbal skills of production and labeling, comprehension, and fluent verbal production and grammatical maturity.

The results provoke some hypotheses regarding the ontogeny of behavior in infancy. It is reasonable to speculate that the child who precociously develops the propensity to manipulate objects and appreciate their contingent perceptual consequences may soon apply this same strategy to the less consistent contingencies of social situations. A natural outgrowth of reciprocal social behavior is imitation, perhaps the parent imitating the child first and then the child imitating the parent. Early imitation will likely involve gross and fine motor behavior, but with development and the presence of a verbally fluent model the imitation of rudimentary language behavior subsequently emerges. Simplistically, the main behavioral trend emphasizes asocial and then social operant learning followed by the propensity to imitate sensorimotor and then verbal behavior. Given adequate verbal models and encouragement, it is perhaps not surprising that a trend involving such basic learning processes should undergird a major developmental trend in infancy and predict later IQ.

*Prediction of childhood IQ*—Does this major developmental trend during infancy, whatever its psychological interpretation, stretch into childhood? Significant correlations between the components of the infant trend with childhood IQ scores do not exist from 6 months, yet the 6-month component is an integral part of the principal developmental trend in infancy. Perhaps it is in this sense that some have claimed that infant tests represent valid indexes of contemporary mental and neurological development, but nevertheless do not predict later IQ.

A second observation is that once verbal items appear on the principal components of the main trend (a few at 12 months, and more at 18 and 24 months), those component scores predict later IQ. Since females display more maturity in vocal-verbal behavior at younger ages than do males, it may not be surprising that the main trend predicts childhood IQ for girls

at 12 months but not until 18 months for boys.

Finally, the predictions for girls from the 12–month principal component to the Binet at 3½, 6, and 10 years are nearly as high as the correlations for the 18–month and 24–month components. In contrast, predictions to the Binet for males are infrequent and considerably lower than for the females. However, the picture is somewhat different when the Wechsler tests are used as the criterion measure of childhood IQ. Sex differences in the prediction of childhood IQ may be dependent on the particular childhood IQ test as well as on the infant test and the nature of the sexes.

*Minor developmental trends*—Although the existence of smaller developmental trends may rest on the significance of a single correlation, it is also the case that minor trends in these data may be imperfect reflections of major behavioral transitions that are actually quite important.

*Males.* One smaller trend for males suggests that there is an inverse relationship, especially for boys, between social orientation and frolicsome play in the first-year test situation and later verbal skills. The importance of this observation derives from the fact that other studies concur in suggesting that males who are either physiologically high-intensity neonates, or who have a relatively strong social and playful nature during the infant test situation, have relatively low standing on verbally oriented assessments later in early childhood. Thus, while early vocal-verbal behaviors have greater predictive salience for girls, early social and frolicsome activity may have inverse predictability, especially for boys.

*Females.* Although the females showed several smaller trends, one pattern stands out because of its comparatively high predictions to later IQ and its relative uniqueness for females. It involves an emphasis on perceptual-verbal facility—one that is distinct from gross and fine motor skills, the imitation of fine motor behavior, and perhaps perceptual-figural (as opposed to verbal) relations. The correlations between components forming this trend and childhood IQ are impressively high (some r's greater than .60) and are relatively unique to females. One very tentative implication is that while the main trend represents advanced performance on a mixture of perceptual-verbal, fine motor, imitation, and perceptual-figural tasks, this smaller independent strain seems to highlight disproportionate skill in perceptual-verbal fluency relative to the other behaviors and emphasizes the special salience that vocal-verbal skills apparently have for females.

**Summary and Conclusions**

Several conclusions can be drawn from this review:

Until the second year of life there is relatively poor prediction from infant tests to IQ assessed in middle or late childhood. Infant tests may have some value in detecting neuromotor abnormalities, however, and a severely low score has somewhat greater predictive significance than average or high scores.

The low predictive correlations are not a function of poor test reliability.

While there are no pronounced sex differences on infant test total scores, girls may show higher correlations with later IQ than boys, especially for some infant tests and for relationships covering early infancy and childhood.

Predictions from 12 to 24 months

may be increased slightly by adding parental socioeconomic class into a multiple-regression formula with infant test score; prior to 12 months, parental socioeconomic status alone is the best single predictor.

Infant tests administered in the first year of life have more predictive power in their item pool than the total score reflects, and early vocal-verbal behavior may have salience in predicting later IQ for girls and frolicsome social activity may have inverse predictability for boys.

Despite the fact that correlations can be increased somewhat by the above procedures, for the most part the level of prediction from the total score for normal children remains modest and of minimum practical utility.

A simple conception of a constant and pervasive g factor is probably not tenable as a model for "mental" development, especially for the infancy period. The data are strong in their denial of simple continuity of general precocity at one age with general precocity at another age during the infancy period, and emphatic in demonstrating marked qualitative shifts in behavioral dispositions. The network of transitions between skills at one age and another is likely more specific and complex than once thought, and not accurately subsumed under one general concept.

---

Condensed with permission from the *American Psychologist*, 27(8):728–748, 1972.
Copyright 1972 by the American Psychological Association, Inc.
Reprints of original article from Robert B. McCall, Ph.D., Senior Scientist and Chief, Perceptual-Cognitive Development Section, Fels Research Institute, Yellow Springs, Ohio 45387

# Comparative Validities of Direct and Indirect Personality Tests

**William A. Scott and Ronald C. Johnson**

Attitudes and other personality constructs may be measured in a variety of ways, including direct self-report, indirect (or disguised) verbal measures, physiological indicators, and the coding of overt behavior or the traces of past behavior. The preference of many psychologists for direct self-report may be traced to at least two sources: theoretical appropriateness and methodological convenience. Theoretical appropriateness applies to constructs, such as attitudes, beliefs, and motives, that are phenomenological in locus; for such constructs the subject may be in the best position to aid valid measurement by his own introspections. The argument for methodological convenience is based on the greater ease with which direct measures can be constructed and scored; it is compelling only to the extent that validity is not greatly sacrificed by the simple method.

The preference of other psychologists for indirect methods of assessment rests on the assumed invalidity of self-report data. This invalidity is attributed to the subject's unwillingness or inability to describe his own true state; it is expected that an indirect method, which circumvents this impediment, will show superior validity.

The most troublesome impediment to research on comparative validity lies in the difficulty of specifying appropriate criteria for the construct that is to be measured. Some criteria—such as overt action, ratings by friends, and membership in groups with known stands on a particular attitudinal issue—are biased in favor of direct measures, since the criteria depend on the person's willingness to display the attribute in question; thus both criterion and predictor are confounded with the same irrelevant source of variance that the indirect measures are designed to circumvent.

Expert judgment should provide a criterion less subject to distortion by the person's style of self-presentation. However, this requires intimate knowledge of the person which would permit the expert to see through any facade that he might present to others. Construct validation—relating scores on the test to theoretically relevant variables—would appear to be an unbiased way of comparing direct and indirect measures. Unfortunately, there are few psychological constructs so well defined theoretically that unambiguous predictions can be made to external variables.

It would seem that the problems of inadequate criteria can be overcome only by comparing validities of two kinds of tests in relation to a wide variety of criteria. If none can be found that is not biased against indirect measures, then it is difficult to see how validity of such tests can be investigated empirically.

The studies reported here concern a range of psychological traits for which meaningful criteria could be found. Five kinds of indirect measures were used: The subjects were required to judge the effectiveness of arguments for and against an attitudinal object; predict the consequences of hypothetical events; write imaginative stories about Thematic Apperception Test (TAT) pictures; complete brief stories by selecting from among alternative endings; and judge the prevalence of certain disapproved orienta-

tion within their peer groups. Two kinds of criteria were used: friends' ratings and expert judgment.

**Study I: Attitudes of College Students**

The subjects were 234 students who filled out three kinds of instruments designed to measure their attitudes toward 14 different issues. The three instruments were self-report, judged effectiveness of arguments, and predicted consequences. Half the subjects took the tests under conditions calculated to enhance the impression of anonymity, while the other half wrote their names on all test booklets and criterion ratings.

The self-report instrument consisted of 56 statements representing 14 different attitudinal issues: Nixon, police, Vietnam war, foreign communism, University of Colorado professors, University of Colorado administrators, sexual promiscuity, homosexuality, legalized marihuana, Students for Democratic Society, miscegenation, parents, Jews and self-denial. The subject indicated his degree of agreement or disagreement with each statement on a 7-point scale. For instance, attitudes toward Nixon were assessed from reactions to the following statements:

Richard Nixon would make a good President of the United States.
The election of Richard Nixon as President would be detrimental to the interests of the United States.
Richard Nixon is one of our great contemporary statesmen.
Richard Nixon is untrustworthy and ill-informed about the problems of this nation.

The instrument for judging the effectiveness of arguments consisted of 56 arguments for the 14 statements. For example:

A good reason for voting for Richard Nixon is that, like Eisenhower, he will end the war without sacrificing this nation's honor.
A good reason for voting against Richard Nixon is that he is incapable of developing social policies that get at the roots of poverty and violence.

For each object there were two pro and two con arguments. The task was to rate on a 7-point scale the effectiveness of each argument for the position indicated.

In the instrument for predicting consequences of events, each event was described in a sentence or two, and the subject predicted its consequences in his own words. ("If Richard Nixon is elected President of the United States, one of his first acts after assuming office will be to . . .."). The subjects made their predictions without regard to their own opinions concerning the desirability or undesirability of the circumstances described.

The 14 issues were selected to represent a range along a public-private dimension. A public issue was one which the subjects would generally regard as legitimately controversial, hence they would see no reason to conceal their true attitudes toward it. A private issue was one on which the subjects would perceive normative pressures toward a single "desirable" answer, hence they would be motivated to conceal opinions. Public attitudes should be quite validly assessed by direct self-report, while private attitudes might be more validly assessed by indirect methods.

Criterion ratings from two friends were obtained on a form that asked the rater to rate the appropriateness of each of 14 statements for describing the subject's attitudes, for example:

I support Richard Nixon for the Presidency of the United States.
I love and admire my parents.

Then the same 14 statements were ranked by the friend according to their relative applicability to the subject, with one representing the "best de-

scription," and 14 representing "least applicable."

Under the identified condition, both the name of the subject and the name of the rater appeared at the top of the instrument. Under the anonymous condition, only the subject's number (designated by himself) appeared.

The self-report instrument was generally insensitive to the manipulated difference in anonymity. Only the attitude toward Jews showed a significant difference between the two conditions in the direction of greater favorableness under anonymity. One interpretation is that Jewish subjects were less reluctant to praise their own group when they were not being identified. Similarly, judgments concerning the effectiveness of arguments were not affected by anonymity.

Predictions of consequences showed several differences. Expressed attitudes toward professors tended to be less favorable under anonymity, while expressed attitudes toward sexual promiscuity, homosexuality, marihuana, and racial intermarriage tended to be more favorable.

From the other significant differences, it might be inferred that for certain issues identified as private (or noncontroversial), anonymous conditions are likely to generate more counternorm responses as one of the indirect measures of attitudes—the prediction of consequences—than are conditions in which subjects can be identified. No significant differences between anonymous and identified conditions appeared for issues identified as public (or controversial), and no comparable difference appeared for any issue on the direct self-report instrument. If one believes that "true attitudes" toward private issues tend to be less socially accepable than those which subjects will admit when they can be identified, these results suggest superior validity of the indirect measure-predicted consequences. However, more evidence on validity is needed, since a socially acceptable response is not necessarily a valid response.

Results based on friends' rankings as criteria were essentially the same as those obtained from friends' ratings, so only the correlations with criterion ratings are reported. There was no evidence that the validity of any instrument depended on the public or private character of the attitude being assessed. There was also no overall effect of anonymity on the validity of any instrument. Considering the five most private issues, the direct measure tended to be more valid than either indirect measure, even under identified conditions, where the reason for distortion was presumably the greatest.

There was a significant increase in mean validity of all three instrument types, from the identified to the anonymous condition, for the most private attitudes. Thus, one may infer that the anonymous condition improved measurement of the relatively private attitudes, regardless of whether these were assessed by direct or indirect means.

The 14 attitudes fell into clusters, which were negatively intercorrelated. The first cluster consisted of: Nixon, police, Vietnam war, University of Colorado administration, parents, self-denial, and (marginally) University of Colorado professors. The second cluster consisted of: foreign communism, sexual promiscuity, homosexuality, marihuana legalization, Students for a Democratic Society, and miscegenation. It seems reasonable to refer to the two clusters as proestablishment and antiestablishment, respectively. Attitudes toward Jews did not go with either cluster.

**Study II: Motives of College Students**

The second study was performed largely by graduate students in a class on attitude assessment. The problem for the six teams of students was to define a personal need or motive, construct both self-report and indirect measures of it, and validate both measures against ratings provided by the subject's friends.

The self-report instrument consisted of 81 statements representing the nine needs: achievement, activity, affiliation, cognition, creativity, dominance (power), contemplation, privacy, and self-punishment.

Eight pictures from the TAT were selected with the expectation that three or more would elicit stories relevant to each of the nine needs. The subjects wrote stories in response to four standard probes—(*a*) What is happening? Who are the persons? (*b*) What has led up to this situation? That is, what has happened in the past? (*c*) What is being thought? What is wanted? By whom? (*d*) What will happen? What will be done?

Friends' ratings were obtained on forms that emphasized the need for accuracy and frankness and that defined each need by a sample of items from the self-report instrument. The friend rated subjects on each need, using a scale ranging from seven ("one of his/her few very most important needs") to one ("completely unimportant to him/her"). Anonymity of raters and subjects was stressed, and the forms were returned in sealed envelopes.

The subjects consisted of 53 students in general psychology who volunteered to participate in a study of "personality assessment." First, eight TAT pictures were administered, followed by self-report measures of the needs. Then the subjects were given two forms to take to close friends who were to rate them on the nine needs, and return the forms in sealed envelopes.

All but one self-report scale attained a validity coefficient significantly greater than zero and the mean was .38. By contrast, only one TAT-based measure correlated significantly with its criterion, and the mean was .05. It is possible that the criterion was biased in favor of the self-report instrument, by explicit listing of some of the items in defining the needs for the friends. However, the criteria for scoring TAT stories were in many cases direct generalizations from the self-report items, so one would expect a modicum of test validity from this source. That it did not generally emerge is evidence that indirect measures of personality traits are inferior to direct measures when friends' ratings are used as criteria.

The nine needs formed two major clusters, one of which included rather passive concerns and the other a more active orientation. The pattern of intercorrelations from the self-report instrument matched fairly well that which obtained among the criterion ratings; the correlation between these two correlation matrices was .71. The matrix of intercorrelations among the motives measured from imaginative stories was different, correlating only .11 with the criterion matrix and .06 with the self-report matrix. Evidently the variables coded from the imaginative stories have different meanings from those assessed by the subject's friends and by the self-report instrument.

**Study III: Counternormative Tendencies of Young Inmates**

Study III was performed to counteract possible biases in favor of self-report instruments. The subjects' personality attributes and circumstances of assessment were chosen to maxi-

mize the likelihood that subjects would be motivated to conceal their own true characteristics. Such conditions should favor indirect methods of assessment. In addition, the criteria chosen for validating the tests were the opinions of "experts" who knew the subjects well and who, presumably, would be less inclined than friends to distort their reports of the subjects' characteristics. The subjects were 92 inmates of a Federal Youth Center who had served at least 6 months in the institution and whose reading comprehension level was 6th grade or higher.

Personality characteristics were chosen for which response norms could be assumed to apply within the institution, hence deviants should be reluctant to answer honestly when questioned directly about themselves. This assumption was verified through informal discussions with inmates following the test sessions. The characteristics were: liking for other inmates, dislike for "rats," intent to be rehabilitated, reluctance to fight, acceptance of blacks, rejection of hippies, respect for officers, job skills, dislike for "con men," rejection of homosexuals, acceptance of Mexicans, long-term goals, appreciation of Federal Youth Center program, respect for adult authority, acceptance of testers, and acceptance of Indians.

The first instrument, Imaginary Situations, was introduced by the following instructions:

> Here are some imaginary stories about situations that might happen in the Federal Youth Center or on the outside. We would like to know how you think each situation would turn out. Please read the circumstances described and think about them. Then, from the four possible outcomes listed, please check the one which seems most likely to you.

There followed 34 situations, two for each of the characteristics. Essentially, they were projective measures of the story-completion type, modified for fixed-alternative responses. For example, liking for other inmates was assessed by the following:

> Jim came into the Federal Youth Center as a new inmate. His first day some of the other guys came up and started asking him about where he came from, what he was in for, and the like. What do you suppose these guys wanted? (Check one.)
> They wanted to make friends to help Jim from getting lonesome.
> They wanted to meet someone new, just because it would be fun.
> They wanted to con Jim into some deal they were planning.
> They wanted to get him off to one side where they could skull drag him.

The second questionnaire, Judgements of Other Inmates' Attitudes, had the following instructions:

> The purpose of this questionnaire is to get your judgment of how most people feel about different controversial matters. The questions do not concern how you personally feel, but what you think are the opinions of those around you. So, regardless of your own personal views, do your best to estimate the attitudes of the people designated in each question.

There followed 102 statements that the respondent was to check "generally true" or "generally false." The attribute "liking for other inmates" was represented by the following items:

> Most of the inmates here like each other and get along with each other pretty well.
> These guys hate each other's guts.
> Most of the inmates trust each other.
> There's hardly an inmate here who trusts any other inmate.
> Most of the guys here would be pretty good friends on the outside.
> Most of the guys in here would not associate with each other on the outside.

The third questionnaire, Study of Inmates' Attitudes, was introduced by the following instructions:

> In this questionnaire, we are interested in finding out your own attitudes toward various people and activities. Please answer the questions in a way that best expresses your own opinions.

This was followed by 102 statements to which the respondent replied by checking "generally agree" or "generally disagree." The first two items for assessing attitudes toward other inmates were:

When you get to know them, most of the guys in this place are pretty nice.
Most of the guys in this place are no good.

No explanation was offered other than the instructions printed on the questionnaires themselves. In general, the testing situation was structured in a manner sufficiently uninformative to permit the suspicion that results might be used against the subjects. This was intended to maximize the chance for superior validity of the indirect instruments.

The criterion measures consisted of ratings of inmates' behaviors and inferred attitudes provided by counselors and work supervisors who knew the subjects well. The rating form presented a set of 7-point scales for rating the subject on each of the attributes and a set of 17 items each presenting several behavior descriptions ordered according to their favorableness with respect to a particular attribute. For instance, attitude toward other inmates was represented by a 7-point rating scale with labeled categories ranging from "very positive" to "very negative," and by the following description:

This man:

Gets along well with the other inmates and is liked by everyone.
Seems to like the other inmates all right, but is not especially popular with them.
Likes some of the inmates, and not others.
Likes very few of the other inmates.
Is a "loner" who dislikes, and is disliked by, everyone.

The results tend to confirm the assumptions about response norms for the psychological attributes. On the instrument entitled Imaginary Situations and on the self-report instrument a significantly large number of the means deviate from the scale midpoint in the normative direction, and the mean deviation for the entire set is more than one-fourth the maximum possible deviation.

By contrast, Judgments of Other Inmates' Attitudes did not show any particular normative effect. Thus, one might infer that when responding for themselves, on either the direct or the indirect measures of these attributes, inmates tended to give answers that corresponded to the investigator's a priori conceptions about what constituted "acceptable" responses in this context. However, when judging their fellow inmates on these same attributes, the subjects reported no general pronorm consensus. Officers' mean ratings yielded a similar result.

The mean correlations with criterion ratings are .14, .08, and .18 respectively, for Imaginary Situations, Others' Attitudes, and Own Attitudes. Only the second and third of these means differ significantly from each other. Thus, in spite of an intentional biasing in favor of the indirect instruments, neither of them showed any consistent superiority over the direct self-report questionnaire, and one appeared generally inferior.

Assuming that criterion ratings provided by the officers constitute valid descriptions of each inmate's characteristics, the intercorrelations among these criterion ratings over inmates may be interpreted as representing "true" interrelations among the traits in this sample. Hence, intercorrelations among the same attributes may be compared with the criterion intercorrelations to see how well that instrument detects the "true" trait interdependencies as seen by the officers. A large cluster, labeled "tractability," has as its central attribute the

judged probability of rehabilitation. Intercorrelations within this cluster tend to be about as high as the scale reliabilities, indicating that officers were not discriminating well among the several components, but tended to respond wth a global judgment about the inmate's performance and prospects.

The cluster, "behavior deviations," represents attitudes toward various kinds of behavior to which the institution generally attached negative sanctions. Officers' judgments of these attributes depended largely on whether or not the inmate was seen as performing the activities himself or associating with other inmates who did. They seem to be judged quite independently; the cluster is established on a priori grounds, rather than on the basis of interattribute correlations, as was the case for the tractability cluster.

The "ethnicity" cluster consists of judged attitudes toward the four major ethnic groups represented in the institution, including whites. Roughly similar intercorrelation matrices were found for the three instruments administered to inmates.

By these considerations, the self-report comes out superior to Imaginary Situations as an instrument for assessing distinct attributes that correspond in meaning to the attributes judged by the "experts" who provided the criterion ratings.

These studies have shown no evidence for the superior validity of any of the indirect measures of personality characteristics over the direct measures against which they were compared. But the conclusions are limited by the criteria—judgments of peers and experts. If comparative validity studies using such criteria are not convincing, it remains to designate procedures that would be. If superior validity of any indirect measure is to be demonstrated, it should probably be sought in construct validation studies, predicting theoretically relevant correlates from both kinds of measures. This will require more dependable predictions about external correlates of the measures than is generally permitted by the current state of theory and empirical knowledge. In the meantime, the use of indirect measures will continue to rest more on preferences of the investigator than on systematic empirical evidence for their superior validity.

---

Condensed with permission from the *Journal of Consulting and Clinical Psychology*, 38(3):301–318, 1972. Copyright 1972 by the American Psychological Association, Inc. Reprints of original article from William A. Scott, Ph.D., Department of Psychology, University of Colorado, Boulder, Colo. 80302

# Getting Information from School and Clinical Psychologists

**Julia R. Vane**

According to the American Psychological Association Code of Ethics,

> test results or other assessment data used for evaluation or classification are communicated to employers, relatives, or other appropriate persons in such a manner as to guard against misinterpretation or misuse. In the usual case, an interpretation of the test result rather than the score is communicated. When test results are communicated directly to parents and students, they are accompanied by adequate interpretive aids or advice.

Under Case 14.B it is further stated,

> the Code does not state that test scores should never be given to a client, but rather places the responsibility for such a procedure on the psychologists. . . . In the last analysis, practice must be determined by . . . good judgment. . . . Psychological jargon can be confusing if not frightening to parents. It is often necessary to write two different types of reports, one intended for parents or teachers, the other for the psychologist or the professional files.

This survey was an attempt to determine the practices of school and clinical psychologists with respect to reporting results of psychological evaluations to the individuals concerned. The survey was conducted by mail, and postmarks were used to determine the area in which the respondents lived. Questionnaires were returned from 137 psychologists from 14 States. Sixty-one percent indicated their primary position was that of a school psychologist. Psychologists in private practice comprised the next largest group, 16 percent of the sample.

Eighty-six percent of those responding rarely or never report intelligence test results in terms of specific numbers to adults or adolescents evaluated or to parents of children evaluated, and 88 percent rarely or never give written reports of psychological evalutions to adults or adolescents evaluated or to parents of children evaluated.

Because so few psychologists indicated they supply written reports, the question which inquired about written reports is not particularly meaningful. An analysis of the responses from the 22 psychologists who said they give written reports almost always or fairly often shows that only 11 of the 137 psychologists replying give the same amount of information in written reports to adults, adolescents, or to parents of children evaluated as they do in written reports to other professionals.

Fifty-one percent indicated that they rarely or never include in verbal reports given to individuals evaluated the same amount of information they would give to professional colleagues. Only on this question did psychologists from New York State differ from those in other parts of the country. Fifty-nine percent of the New Yorkers rarely or never include information compared to only 38 percent from other parts of the United States.

In order to determine whether the type of setting in which psychologists work influences their behavior in reporting, the 50 psychologists who worked only in schools were compared with the 48 psychologists who worked full-time in private practice or who combined private practice with another position. Significantly more psychologists in private practice almost always or fairly often reported intelligence test results in terms of specific numbers to adults, adolescents, and parents of children evaluated. Significantly more psychologists in private practice almost always or fairly often gave written reports to adults,

adolescents, or parents of children. The results also showed that 42 percent of the school psychologists and 45 percent of the psychologists in private practice would include in verbal reports to the individuals involved the same amount of information they would give to other professionals.

On the whole, the results suggest that many psychologists have interpreted the Code of Ethics to mean that results, and particulary specific IQs, should be withheld from individuals involved.

Many of those who answered expressed concern that giving results to parents and individuals evaluated will lead to misuse, misinterpretation, and distress. But with comparisons of any sort, whether it be in the selection of athletes, beauty queens, or merit scholarship winners, the means of selection will be misinterpreted, misused, and upsetting to some involved. Although the desire to avoid misinterpretation of the results is a laudable one, the practice seems to imply a basic distrust of those who come to us for help. It suggests that we and our colleagues are so superior that we must determine what is good for the individual rather than permitting the individual to use the results we obtain to determine his own welfare.

Experience with parents of children who are brought to Hofstra Psychological Evaluation Center, where we do give reports and IQs complete with explanatory charts, shows that many of them feel bitter about previous experiences with psychologists who denied them reports when requested. Many parents state they will not come if they do not get a report. Others, assuming reports will not be given, come to the conferences armed with tape recorders or stenographic notebooks. Some have told us they had obtained reports through the ruse of having the report sent to the family physician who then turned it over to them.

Does the psychologist in his attempt to protect the individual from himself really help the individual, or does he give rise to a different sort of interpretation? An example of the upset that withholding information can generate was demonstrated by an education student who said she had her child tested by a psychologist, who, after verbally interpreting the results, refused to give her a report or specific IQ. He offered to send the results to the child's teacher. When the parent pointed out that she was a teacher too, the psychologist replied he could not give her the results because she would become too upset. Whether the psychologist said this I can't be certain, but I do know that the woman was much more upset by the withholding of information than she would have been by whatever the report would include.

What does this kind of behavior do to the field of psychology in the eyes of the public? Is it sensible to withhold information about the IQ when the publishing field is flooded with books on how to raise your child's IQ? It would appear that the time has come for psychologists to consider a change in their behavior with respect to reporting results of psychological evaluations to the individuals involved.

---

Condensed with permission from *Professional Psychology*, 3(3):206–208, 1972. Copyright 1972 by the American Psychological Association, Inc. Reprints of original article from Julia R. Vane, Ph.D., Professor of Psychology, Hofstra University, Hempstead, N.Y. 11551

# mental health and the law

## Courts, State Hospitals, and the Right to Treatment

**Jonas Robitscher**

During the 1960's, when community psychiatry was being developed, the State hospital patient continued to suffer from neglect. In the 1970's a series of legal decisions has created a surge of interest in the State hospital patient by enunciating a doctrine of judicial responsibility for the maintenance of standards.

Although most fair-minded people would applaud the overdue recognition of the need for a more therapeutic approach to the State hospital patient, court intervention into the scene raises questions as to the appropriateness of judicial supervision of psychiatric institutions. It marks a new era in psychiatry when psychiatrists must do their institutional reforming and their housecleaning in public, under the scrutiny of the court, and when they must conform their practices to court-approved standards.

The cases revolve around the concept of the right to treatment. This concept dates back to 1960, when the term was used in an article in the *American Bar Association Journal*. Right to treatment is a shorthand expression for much more than a statement about health care delivery; it means the right of involuntarily hospitalized mental patients to receive adequate therapy in exchange for being deprived of liberty.

Birnbaum, the physician-lawyer who originated the concept, was not advocating psychiatric treatment as a right for all. He addressed himself solely to the involuntarily hospitalized State hospital patient who was not receiving the care available to private psychiatric patients and who presumably was being held for a longer period than those patients.

His proposal suffered from neglect; then it was applied to cases that fell outside the scope of his argument; now it is being applied, among other examples, by a Federal judge to the Alabama State hospital system. But in the process the court is becoming involved in the internal administrative affairs of State hospitals in a more all-encompassing way than Birnbaum proposed.

### Legal and Political Aspects

Birnbaum's concept was both legal and political. Involuntary mental hospitalization is a deprivation of liberty. The Fifth Amendment prohibits the deprivation of liberty without due process of law. The benefit that accrues to patients who get adequate psychiatric treatment enables them to regain their liberty at the earliest possible time and so gives legal legitimacy to the hospitalization. But if hospitalization does not provide the care that would promote recovery, the due process standard is not met and the patient should be released even if he represents a potential danger to himself or to others. The concept that not even dangerousness should be the excuse for deprivation of liberty has both civil libertarian and political implications. Freeing inadequately treated patients would represent justice to the patient; it is also designed to put pressure on State legislatures to appropriate funds for pro-

per treatment programs in order that dangerous patients will not have to be freed.

The heart of Birnbaum's proposal is to use simple yardsticks—the one he originally relied on was the ratio of personnel to patients—to gauge adequacy of treatment. Birnbaum is a lawyer and not a psychiatrist, but he is a physician and identifies with the therapist; he has emphasized that courts are not the suitable agent to consider the details of a treatment plan or to determine appropriateness of therapy.

The enforcement of the right to treatment under the Birnbaum proposal would be by a habeas corpus proceeding. By demonstrating that hospitals were staffed below a minimum ratio or that consultations were spaced farther than minimal periods of time, the patient would be entitled to release. Recently, expanding his pragmatic standards for proof of adequacy of treatment, Birnbaum spelled out a more complete list of desirable institution-wide standards that must be met for an inmate to be retained: accreditation by the Joint Commission on Accreditation of Hospitals; qualification for Medicare and Medicaid funds; physical standards that meet minimums set by the American Psychiatric Association; personnel-patient ratios that meet the standards for private hospitals that were set by APA and were used until 1969, when patient-personnel ratios were discarded as "meaningless as a general standard"; State licensing of all professional hospital personnel; regular progress notes reflecting patient-physician consultation; and presence of halfway houses and other intermediate facilities.

The cooperation of State legislatures is needed in order to have effective treatment. Some courts may not go as far as Birnbaum might wish in enunciating this right. The legal process is time-consuming. For all of these reasons, Birnbaum has also advocated a second approach to the right of treatment by State legislative action. A bill to specify this right has been introduced into the State legislature of Pennsylvania every year since 1967; one version is now under consideration.

**Important Court Decisions**

The first reaction to Birnbaum's proposal was a prompt endorsement in the *American Bar Association Journal;* but little more was heard of it until it was seized upon by Judge David Bazelon 6 years later in a District of Columbia case involving an offender who had pled not guilty by reason of insanity when charged with a minor offense and had been held in St. Elizabeths Hospital far in excess of the maximum criminal penalty.

The position of the offender who is diverted from the criminal process and committed to a mental hospital raises constitutional questions, but this is not the type of situation with which Birnbaum had been concerned. On the one hand, the criminal offender is in a better position than the average mental hospital patient to claim an abridgment of freedom when he is held an indeterminate period; he often does not meet the criteria of mental illness (the appellant in Judge Bazelon's case was diagnosed as a "sociopath") and he can argue that criminal justice procedural safeguards are not being met. The argument can also be advanced that there is no recognized psychiatric treatment for sociopathy and that hospitalization until the person is "cured" can represent a life term. In contrast, the average involuntary State hospital patient can be seen as being deprived of

his liberty without having infringed any law; he has no recourse to a strict system of procedural safeguards to guarantee that the deprivation of liberty will be appropriate, effective, and for no longer than a minimum period. This is the patient for whom Birnbaum devised his legal remedy.

The *Rouse* case, in which the right to treatment was first judicially noted, applied the Birnbaum concept only to a class of patients being held for purposes other than merely restoration of mental competency; it did not deal with the average committed patient not under criminal charges. The case disappointed Birnbaum because it failed to adopt the idea of rough yardsticks that would automatically determine if appropriate treatment were available.

Instead of trying to enunciate such a yardstick as patient-personnel ratios, the court sent the case back to the lower court for a full hearing on the adequacy of treatment. This was the Pandora's box which Birnbaum had said should not be opened; courts do not have the time or the competence to decide in individual cases if patients should have one type of treatment in preference to another. The *Rouse* rehearing, which is transcribed in a record of more than 500 pages, required the attention of the whole court for varying parts of 4 different days: it required the time of a United States District Judge, three lawyers for Rouse, two lawyers for St. Elizabeths, the director of the hospital, a staff psychiatrist, a staff clinical psychologist, two ward attendants, a psychiatrist on the staff of the District of Columbia Commission on Mental Health, and two psychiatrists to serve as expert witnesses for Rouse.

Although the *Rouse* case clearly said there was a right to treatment, the basis for this right was said to be a District of Columbia statute; the question of the applicability of a Fifth Amendment right went unanswered.

Additional cases in the District of Columbia and Massachusetts affirmed the concept of the right to treatment, but until the recent case of *Wyatt* v. *Stickney* no case affirmed this as a constitutional right and as a right for all (not merely quasi-criminal) patients. In one Pennsylvania case, the court expressly declined to consider the right when this was demanded for the inmates of the State hospital for the criminally insane because then the court "quite possibly would have to supervise the therapy of several hundred persons."

Two recent Bazelon decisions have defined further how far the courts will involve themselves with the appropriateness of the treatment of quasi-criminals confined to mental institutions. Conscious of the quagmire that judicial review of hospital administration could enter, Bazelon apparently is narrowing the role of the court. In *Tribby* v. *Cameron* it is stated that it is not the function of the court to substitute its own judgment for the judgment of the hospital administrators, but it is a function of the court to see that the administrators have made permissible decisions. A permissible decision is one that takes account of the relevant information which was available or should have been available to the decisionmaker. *Covington* v. *Cameron* further defined the limited judicial review that Bazelon favors; the court must assure itself that the decisionmakers "have: 1) reached a reasoned and not unreasonable decision, 2) by employing the proper criteria, and 3) without overlooking anything of substantial relevance."

Such decisions by the courts are part of a movement that became obvious in the late 60's of court concern

in areas where the person operates at a disadvantage because of the power of the system, such as the administration of corrections, the juvenile justice system, and public welfare law.

In two decisions, courts have examined the internal affairs of mental hospitals, found much that is wanting, and ordered the hospital to remedy conditions. One case does not take us too far from the *Rouse* case; like most cases in this area, it dealt with a quasi-criminal and the imposition of an indeterminate hospital commitment in place of a jail sentence. *McCray* v. *Maryland* involved the Patuxent Institution, an intermediate institution with features of both a hospital and a prison. The court issued 35 pages of rules and regulations covering all phases of the administration of the institution, down to the details of the personal possessions the inmate can have in his cell and the edibles he can possess. The main body of the 46-page decision considers the treatment available at Patuxent. The court found that sections of the institution used for solitary and disciplinary confinement were not treatment facilities (although the psychiatric staff stated that these were used as "negative reinforcers" in a therapy program). The court found that the program's aim of a total approach to rehabilitation was not being met by the institution's program.

### The Alabama Case

The one case that embraces Birnbaum's right to treatment theory by applying it to all involuntary patients (not merely those who because of law violations might be given extra criminal law procedural safeguards) and by asserting that this is indeed a Fifth Amendment right is *Wyatt* v. *Stickney,* a 1971 Federal Court decision concerning Bryce Hospital. The *Wyatt* case decision ordered the institution, its superintendent, and the Commission of Mental Health for the State to mount an effective treatment plan; it retained jurisdiction over the case so the court could see that this was done. The two other Alabama State hospitals, Searcy and Partlow School for the Retarded, were included in later aspects of the case.

In an order 6 months later, Judge Johnson specified the three fundamental conditions that his court would examine to ascertain if adequate and effective treatment was available. These were: a humane psychological and physical environment; qualified staff in numbers sufficient to administer adequate treatment; and individualized treatment plans. The court was aware of the quagmire it would be involved in if it considered the appropriateness of treatment plans; it stated it would require only that there be individual plans.

The court noted that the dormitories were barnlike structures with no privacy for the patients; for most patients there was not even space that a person could think of as his own. The toilets seldom had partitions between them; wearing apparel was shoddy; and work assigned to patients was nontherapeutic. Only 50 cents per patient per day was spent for food. (In Maryland, some patients have been diagnosed as having pellagra, and diets in the State's facilities for the mentally ill and mentally retarded have been found to have been deficient in protein, iron, and calcium.)

Concerning a qualified and numerically sufficient staff, the Alabama court stated that "more psychiatrists, Doctor of Philosophy level psychologists, and qualified Medical Doctors are not only a medical but are also a constitutional necessity in this public institution."

Responding to Judge Johnson's order, the State of Alabama's Department of Mental Health has submitted a budget more than double that of previous years. It would provide for a patient-staff ratio of better than one-to-one, a per diem per patient expenditure of more than $25, and an increase in hospital pay scales to match those of the Veterans Administration.

**Standard Setting by the Court**

The court has not only decreed that psychiatrists, psychologists, and social workers must meet the same licensing and certification requirements as are met by persons of the same profession in private practice, but has also set minimum staffing for every 250 patients: one unit director, two psychiatrists, four physicians, 12 registered nurses, six licensed practical nurses, 92 aides and 10 orderlies, one Ph.D. psychologist, one M.A. psychologist, two B.S. psychologists, two M.S.W. social workers, and five B.A. social workers. Requirements for other health personnel, clerks, typists, and a chaplain (one-half chaplain for every 250 patients) bring the total to 207.5 personnel for 250 patients, a 1 to 1.25 ratio. The requirement of two psychiatrists for every 250 patients will mean that a State system that employed only four psychiatrists will now have to employ 42.

The Alabama case has set a train of other cases in motion. In Georgia the American Civil Liberties Union has filed a suit alleging patients receive inadequate diagnosis, care, and treatment. *The New York Times* has noted that institutions for the retarded in New York and Massachusetts are the objects of pending suits and that several national mental health organizations and legal associations have said they plan to file similar suits in many other states.

The *Wyatt* case ruling goes far in creating a legal right to the kind of mental hospital treatment that meets court-determined standards of care. This agrees with the policy of groups that have supported the court proposals in "friend of the court" briefs; these have included the American Psychological Association, National Legal Aid and Defenders Association, American Orthopsychiatric Association, American Civil Liberties Union, and the Center for the Law and Social Policy. The last three have formed a National Council on the Rights of the Mentally Impaired to bring similar suits in other States.

The case raises many questions, including the matter of logistics. Will well-qualified mental health personnel be found to fill the jobs that the court has ordered created, or will the discrepancy between the incomes of public and private practitioners and the differences in working conditions continue to shunt less well-qualified practitioners into the State system?

Another question concerns the relationship of therapists to other personnel. The 250 patients will be required to have the services of 207.5 personnel but only two will be psychiatrists. Would a different kind of personnel structure, emphasizing therapy by psychiatrists and giving less emphasis to other medical personnel as well as to custodial and ancillary personnel, give the institution a different kind of atmosphere?

This leads to a further question: Should money be spent to rebuild existing institutions, or should money be spent to find alternatives to the State system? Perhaps the two approaches are not mutually exclusive. An American Civil Liberties Union attorney has said that States will probably find it less expensive to provide alternate facilities and that the effect of the

ruling will be to release many patients to facilities such as halfway houses. Such facilities would eventually have to face court scrutiny to determine if they were providing adequate care, but since they would be caring for voluntary patients this would be legally less compelling—but not a socially less important—issue.

**A Suggested New Approach**

In a more detailed consideration of the right to treatment, I have outlined a program that would recognize the right to treatment but would see treatment not as an automatic consequence of improved patient-staff ratios or greater per patient expenditures, but as a result of a changed attitude of psychiatry toward the State system. The changed attitude would have to include greater interest by the "private sector" of psychiatry—university programs and private practitioners; it would have to lead to the phasing out of the dual system of patient care with different standards for public and private hospitals and would require greater responsibility by psychiatrists for the determination of objective standards of patient care so that the quality of the product delivered by psychiatry can be measured.

As difficult as the measurement of good psychiatric care may be, some standards will have to be set by psychiatry so that patients, third party payees, State legislatures, and courts can compare hospital A and hospital B, treatment A and treatment B, psychiatrist A and psychiatrist B. The myth that all psychiatric care is equivalent to all other psychiatric care must be abandoned, and realistic methods of appraisal of care must be substituted.

Innovative approaches to hospitalization will be needed before patients receive adequate treatment. Only an approach more comprehensive and more sophisticated than the *Wyatt* case approach will succeed in upgrading the State hospital system.

We can debate whether the *Wyatt* case is progressive or regressive. It provides justice for patients, but brings the courts into the mental health movement in a way that may turn out to be harmful for both court and hospitals.

Psychiatrists must recognize that there is a right to treatment. We can also insist that courts use simple yardsticks—which psychiatry will have to develop quickly—to keep the legal determination of treatment adequacy from usurping psychiatric authority and determining hospital policy.

---

Condensed with permission from the *American Journal of Psychiatry*, 129(3):298-303, 1972. Copyright 1972 by the American Psychiatric Association. Reprints of original article from Jonas Robitscher, J.D., M.D., Henry R. Luce Professor of Law and the Behavioral Sciences, Emory University Schools of Law and Medicine, Atlanta, Ga. 30322

# mental health news digests

### Study of "Labeling" of Children

Many children today are "labeled," or diagnosed, or classified, or described with terms such as "mentally retarded," "juvenile delinquent," "emotionally disturbed," "culturally deprived," and so on. What is the effect on a child who is tagged with one of these labels? Are there dangerous and permanent consequences? Can diagnostic procedures be improved? What are the policy implications and even the possibilities for legislative action?

A project, sponsored by five agencies of DHEW, is now underway that will attempt to deal with the many problems related to labeling of children. Dr. Nicholas Hobbs of Vanderbilt University is directing the study. Information about it may be obtained from Martin Goula, Office of Child Development, Box 1182, Washington, D.C. 20013.

### Teaching Children How To Read

Children with reading difficulties may profit from a linguistics study now being conducted. In one phase of the project, the investigators plan to improve and expand a syllabary method of teaching reading to youngsters as an aid to more advanced reading.

The syllabary method uses the syllable as the basic language unit rather than the usual phoneme, which breaks the sound and structure of words into letter-by-letter components. In a reading primer based on the syllabary approach, some syllables may be represented as pictures. For example, a drawing of a bumblebee may indicate "be," as in the first syllable of "before." Thus the relation between the sound and meaning is obvious to the child.

The investigators plan to use a combination of pictures and simple written syllables to construct the stories. "Reading" of these materials can be accomplished almost immediately, satisfying the child and motivating him to further reading development. Gradually he learns that the syllabic units may be combined to yield additional meaningful words. In this manner, he is shown that orthography—the written word—parallels the sounds and meaning of language. When the child is competent in the syllabic method of reading, he is taught that the alphabetic unit—the phoneme—is a more efficient means of word production than is an increasingly large number of memorized syllables and pictures.

The project also calls for the investigation of several related areas of linguistic acquisition. Children's "linguistic environments" will be analyzed by comparing the manner in which parents speak to their own children with the manner that adults generally use in speaking to young children. This study of parent-child dialogs will be helpful in understanding how children acquire language patterns.

The investigators are Dr. Lila R. Gleitman, professor of linguistics in the School of Education and Dr. Paul Rozin, professor of psychology, at the University of Pennsylvania.

### Relationships Between Children's Personalities and their Cognitive Abilities

Three groups of preschool children selected from day-care centers and nursery schools will participate in a study of the relationships between personality and learning difficulties. Among the groups to be studied are hyperactive children and withdrawn children, as well as children who show neither extreme in personality. Personality types will be assessed on the basis of both classroom observation and performance in a cognitive situation.

The study will attempt to determine whether or not different types of cognitive errors—for example, irrelevant responses or silence—can be related to the different personality types. The results are seen as particularly relevant to the understanding of learning disabilities and to the design of special education programs for children experiencing these handicaps.

Dr. Marion S. Blank of the Albert Einstein College of Medicine in the Bronx, New York, is principal investigator.

### Workshop on Psychobiological Mechanisms of Electroconvulsive Therapy

The use and effects of electroconvulsive therapy were the subjects of an international conference held recently for top authorities on the treatment of the mentally ill. The conference was sponsored jointly by the New York Medical College and the National Institute of Mental Health.

More refined and safer procedures have rekindled the interest of mental health professionals in this kind of therapy, which was first introduced in 1935. The objectives of the meeting were to stimulate discussion among scientists and, through publication of the proceedings, to provide the public with a clearer view of the issues and problems that surround ECT research.

Psychological and biochemical studies of ECT in man and animals have provided scientific information regarding the clinical effects of induced seizures as well as their effects on memory. The conference dealt with the psychobiology of electroconvulsive treatment and recently reported findings. Current psychobiological studies promise a greater understanding of the therapeutic and generalized effects of ECT on psychological functioning in humans.

Since its introduction, ECT has been used in the treatment of many psychiatric disorders. Studies generally have shown it to be a more effective therapy than antidepressant medications in the treatment of severe depressive illnesses.

The 27 scientists attending the workshop participated in panels on clinical efficacy, biochemical mechanisms, and behavioral effects of ECT; a fourth panel summarized the discussions and suggested areas for future research. Coordinators of the Conference were Dr. Max Fink, executive director of the International Association of Psychiatric Research, Dr. Seymour Kety of Harvard University, Dr. James McGaugh of the University of California, and Dr. Steven Secunda of the National Institute of Mental Health. Inquiries may be directed to Dr. Secunda.

### Study of Rehospitalization

Nearly one-third of all former mental patients return to the hospital within a year of their discharge. A research project now underway is trying to learn

why so many patients return to the hospital and what can be done to help them adjust successfully to community life.

The researchers will work with a group of 200 schizophrenic patients and their spouses or parents, and will study the interactions between the recently discharged patients and their families. They will measure the patients' and families' tolerance for symptoms of mental illness. They also will seek to determine personal attitudes concerning the acceptability and usefulness of hospitalization in the treatment of mental illness.

The major significance of the study lies in its "ecological" approach. The investigators will learn, for example, if family members expect the household routine to be changed by a patient's return home. Also, they will measure the level of general stress within a family at the time of a discharge. Such information will indicate whether or not factors operating within the family have an effect on the process leading to rehospitalization.

The Project Director is Dr. Dolores E. Kriesman, Associate Director of the New York State Department of Community Research, New York, N.Y.

### Training Advocates in Child Mental Health

Two schools of social work are now strengthening their training programs in child mental health.

The Portland State University School of Social Work at Portland, Ore., will train social work students in the mental health aspects of day-care planning for young children. The program will also focus on training students to act as professional advocates on behalf of the mental health of the preschool child. The students work in a variety of day-care settings including private day-care centers, low-income centers in ghetto areas, and outreach day-care services, and explore the social worker's role in delivering such services. The program emphasizes the preventive rather than problem-focused aspects of mental health care for young children, the continuity of mental health training on a growth continuum from prenatal through adolescence, and the focus on developing unique child advocacy curriculum material. Dean Gordon Hearn is program director.

The School of Social Work of the University of Wisconsin at Madison will train graduate social workers in child advocacy and community mental health programs. The students will be oriented to three levels of services: institutional treatment in State and local institutions; outpatient programs at local mental health clinics; and outreach programs to involve parents and others in disadvantaged groups in the planning, development, and use of mental health services at outpatient and inpatient levels. The program not only will enable social workers better to aid young children with mental problems, but will train them to identify and utilize effective planning processes. The project is headed by Dr. Martin B. Loeb, director of the School of Social Work.

### Training American Indians in Social Work

Five schools of social work are launching training programs geared to the needs of American Indian communities. Members of the communities involved played an important part in the planning of each of these innovative programs.

The Arizona State University, Tempe, Ariz. in a consortium with the Phoenix Indian School, plans improvement in Indian education in boarding schools and in the training of social workers to give professonal leadership in programs to serve American Indians and others residing in the Rocky Mountain Southwest region.

The School of Social Work at Barry College of Miami Shores, Fla., aims to develop leadership potential among Seminole and other American Indians living on or near three Indian reservations in Glades, Hendry, and Broward Counties of Southeast Florida. The school will encourage and recruit Seminole high school students to attend college, and will recruit and train American Indians and non-Indians as graduate students of social work to deal with unique problems of the American Indian.

The School of Social Work of the University of Oklahoma will recruit, educate, and support American Indians as social workers dedicated to service and involvement of American Indians. The setting for this field experience will be in three towns in eastern Oklahoma and the surrounding counties where there is a high number of Cherokee Indians in small groups or in small towns.

By employing Indian faculty and systematically recruiting Indian and native students, the Portland (Oregon) State University seeks to provide increased educational and career opportunities for American Indians and native Alaskans in the field of social work. The Portland State University of Social Work also plans to acquaint the non-Indian student body, faculty, and social work community with Indian culture, tradition, and history.

The School of Social Work at the University of Denver has a program designed to enhance social work educational opportunities for various cultural groups in this region often excluded from higher education. These groups are the Mexican-American community, the Black community, and the American-Indian community. The ultimate goal of this program is the establishment of three teaching-learning centers, which will provide opportunities for students and faculty members of the school to participate with neighborhood residents in a joint educational effort.

### New Careers Training Program

The first awards under the NIMH's recently launched New Careers Training Program, which is designed to stimulate and support special programs to train new kinds of workers in new ways, have been announced. The new careerists will be trained to function in roles including advocacy, assessment of consumer satisfaction, outreach, and community organization around mental health issues.

Recipients of the twelve awards are:

Department of Psychiatry, University of Washington School of Medicine
Latin-American Youth Center, Inc., Chicago, Ill.
Wheeler Clinic, Inc., Plainville, Conn.
The Thresholds, Chicago, Ill.
North-Coast Opportunities, Ukiah, Calif.
Boston University, Boston, Mass.
Mexican-American Unity Council, San Antonio, Tex.
University of California, Riverside, Calif.
Orange County Community Mental Health Service, Santa Ana, Calif.
Lutheran Deaconess Home and Hospital, Minneapolis, Minn.
Institute for Health Research, Berkeley, Calif.

More information about the program may be obtained from the Division of Manpower and Training, NIMH.

### Training Minority Groups for Careers in Psychology

Opportunities for members of minority groups to prepare for careers in psychology have been enlarged as a result of eight training grants recently awarded to psychology departments and other institutions by the NIMH. The grants will initiate new programs, increase the number of stipends available for minority students, and strengthen existing training programs. The funds will go to programs at the undergraduate, associate of art's, master's, and doctorate levels for training in delivery of services and research in community-oriented programs.

Sacramento State College will have a pilot program to train psychologists at the B.A. level, designed to encourage minority groups to major in psychology.

The California School of Professional Psychology in Los Angeles will train students at the A. A. level as paraprofessionals in psychology, with participants recruited primarily from disadvantaged minority groups.

Meharry Medical College and Fisk University in Nashville have formed a consortium with three private educational institutions, Vanderbilt University, George Peabody College, and Scaritt College; the consortium's activities in psychology will be supported by the grants.

The University of California at Riverside will offer training in community-oriented clinical psychology, with Mexican-American students the primary participants in the program.

San Francisco State College will train practitioners at the master's level in clinical and school psychology. Besides recruiting primarily minority participants, the program aims to attract older students with prior occupational experience.

The Southern Regional Education Board of Atlanta received a grant to develop guidelines for programs in community-oriented clinical psychology. The grant will also enable the Board to assist Black colleges in the South in carrying out their programs.

### Asthmatic Children Helping Themselves

Young children suffering from bronchial asthma may be taught to ward off the suffocating attacks by voluntarily expanding the size of the tubes through which air flows to their lungs the moment breathing becomes difficult.

Recent research has indicated that while an initial asthma attack may have a physiological root, children often associate the attack with emotional or situational events occurring around the same time. Within a short time, social-psychological events such as change of environment or emotional excitement may trigger the majority of attacks the children suffer. These "causes" have proven difficult to extinguish both because they are vaguely defined and also, the attention a child receives during an attack may reinforce his dependence upon the triggering event.

In an attempt to sidestep these barriers, the investigator hopes to train asth-

matic children to control the constriction and dilation of the bronchial tubes into the lungs. Carefully screened young asthma patients whose lung and bronchial tissue has not yet been damaged by the illness will participate in the project. They will be conditioned, or trained, to control the rate of air flow into the lungs under conditions which induce their asthma. Aided by encouragement and reassurance, the children will breathe into a specially designed instrument that will register a light and inform the child that he has been able to increase his peak flow of air.

In early stages of the training, medication will be used if necessary to assist the children to dilate their bronchial passages. Gradually the dosage will be decreased until the children can be told they are successfully producing bronchodilation on their own. At this point they will be asked to practice the new skill as "homework."

The conditoning procedure will allow paramedical personnel to give the treatment and will encourage the skills of nonprofessionals in the delivery of health care.

Dr. Aman U. Khan, child psychiatrist at the Children's Memorial Hospital in Chicago, is conducting the research.

**New Publication**

*The School Psychology Digest* is the new quarterly magazine of the National Association of School Psychologists. Its purpose is to identify and clarify known principles of the behavioral sciences and to assess new ideas generated in the fields of developmental, experimental, and social psychology applicable to preschool, adolescent, and exceptional children. Original, condensed, and abstracted articles will be included, as well as a book review. The subscription rate is $10 a year and can be obtained from Dr. John Guidubaldi, Editor, 311 Education Building, Kent State University, Kent, Ohio 44242.